JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9243–9254, APRIL 16, 2000

Human-activity-enhanced formation of organic aerosols by biogenic hydrocarbon oxidation

Maria Kanakidou and Kostas Tsigaridis

Environmental Chemical Processes Laboratory, Department of Chemistry, University of Crete, Heraklion, Greece

Frank J. Dentener

Institute for Marine and Atmospheric Research Utrecht, University of Utrecht, Utrecht, Netherlands

Paul J. Crutzen

Atmospheric Chemistry Division, Max Planck Institute for Chemistry, Mainz, Germany

Abstract. Tropospheric aerosol can affect climate and the chemistry of the atmosphere. Organic particulates form a significant fraction of the atmospheric suspended matter over forested areas and may originate to a large extent from the oxidation of natural hydrocarbons. A three-dimensional global model of the troposphere is used to evaluate the contribution to the global organic aerosol (OA) source of the secondary organic aerosol (SOA) derived from the ozonolysis of biogenic volatile organic compounds (BVOC) and its evolution since preindustrial times. BVOC have been represented by a mixture of α– and β–pinenes, and their aerosol-forming parameters and chemical reactivities versus O_3, OH, and NO_3 have been estimated using laboratory information. An important factor in SOA formation is the deposition of condensable aerosol oxidation products onto preexisting organic aerosol, and this has been taken into account. The thus-calculated source of SOA is evaluated to have increased from 17-28 Tg/yr in preindustrial times to 61-79 Tg/yr at present. This threefold to fourfold enhancement of the formation of organic aerosol from natural BVOC is attributed to an increase in ozone and organic aerosol from anthropogenic sources. The main uncertainties involved in our calculations are related to the composition of BVOC emissions and the details of their aerosol formation capabilities.

1. Introduction

During the past decade much attention has been paid to the impact of aerosols on atmospheric chemistry and climate and to the ability of hydrophilic aerosols to act as cloud condensation nuclei [*Novakov and Penner*, 1993; *Novakov and Corrigan*, 1996]. At continental midlatitudes, organic material contributes ~20-50% of the total fine particulate mass [*Saxena and Hildemann*, 1996]. In tropical forested areas this contribution can be as high as 90% [*Andreae and Crutzen*, 1997; *Talbot et al.*, 1988, 1990; *Artaxo et al.*, 1988, 1990]. A substantial fraction of the organic component of atmospheric particles was shown to consist of water-soluble, possibly multifunctional compounds [*Saxena and Hildemann*, 1996]. *Andreae and Crutzen* [1997] have suggested that the production of organic aerosols from the oxidation of biogenic organic compounds might range from 30 to 270 Tg/yr and thus can be comparable in magnitude to the production of biogenic and anthropogenic sulfate aerosols [*Pandis et al.*, 1992]. No investigations exist up to date on the possible evolution of this secondary source of organic aerosols since preindustrial times.

The impact of human activities on secondary organic aerosol formation due to the increased levels of the oxidants (ozone, nitrate, and hydroxyl radicals) deserves investigation, since secondary organic aerosols (SOA) might have constituted a major component of condensation nuclei (CN) and cloud condensation nuclei (CCN) over the continents during the preindustrial era. Laboratory studies [*Novakov and Penner*, 1993] have shown that the water-soluble organic species in smoke particles from smoldering biomass combustion are responsible for the CCN activity of these particles. Little is known about the physical and chemical properties of the biogenic SOA. *Kerminen* [1999] calculated that organics with low volatility and high gas phase production rate can contribute significantly to the growth of nuclei to CCN size. *Cruz and Pandis* [1997] observed that dyoctylphthalate aerosol as large as 0.15 μm in diameter did not become activated even at supersaturation as high as 1.2%, whereas adipic and glutaric acid containing SOA can serve as CCN at supersaturations of 0.3% and 1.0%. Their results indicate that CCN activity of hygroscopic organic aerosol may be compared to that of some inorganic aerosols. In contrast, *Virkkula et al.* [1999] found pure organic particles resulting from terpene oxidation to be only slightly hygroscopic. The hygroscopic properties of mixed organic and mixed organic/inorganic particles appear to be even less well established. *Saxena et al.* [1995] concluded that for nonurban locations the presence of organics in atmospheric particles enhances water absorption by inorganics whereas for urban locations it has the opposite effect. More experimental investigations on the chemical composition and hygroscopicity of SOA are required to determine their potential role as CCN.

Copyright 2000 by the American Geophysical Union.

Paper number 1999JD901148
0148-0227/00/1999JD901148$09.00

Laboratory studies have shown that oxidation of several biogenic compounds like α- and β-pinene by OH, O_3, and NO_3 radicals produces organic substances which can condense into particulate matter [*Odum et al.*, 1996; *Hoffmann et al*, 1997; *Christoffersen et al.*, 1998; *Griffin et al.*, 1999a]. Isoprene, the biogenic hydrocarbon with the largest emissions, has a major impact on the global levels of oxidants via gas phase reactions in the troposphere [*Houweling et al*, 1998; *Wang et al.*, 1998; *Kanakidou and Crutzen*, 1999; *Poisson et al*, 2000]. However, its role in organic particulate matter formation is minor compared to the monoterpenes [*Pandis et al*, 1992].

Experimental chamber [*Christoffersen et al.*, 1998] and field [*Kavouras et al*, 1998, 1999] studies have focused on the identification and quantification of the condensable oxidation products of terpene oxidation reactions and the understanding of the chemical and physical mechanisms which lead to this condensation. Recently, some pinene oxidation products were identified and quantified in fine atmospheric particles over forested areas in southern Europe [*Kavouras et al*, 1998, 1999]. These observations indicate that organic acids formed by pinene oxidation might be linked to new particle formation over forested areas since their concentrations maximized simultaneously with that of Aitken particles, whereas sulfate aerosol did not show the same pattern. Chamber experiments provided insight into the process of new particle formation by heteromolecular homogeneous nucleation [*Hoffmann et al*. 1998]. Stable binary diacid adducts, which could form new particles, have been identified in the particulate phase resulting from α-pinene oxidation by O_3 [*Hoffmann et al*, 1998].

The efficiency of particulate matter generation from the oxidation of gaseous biogenic hydrocarbons depends on the chemical structure of the parent hydrocarbon reacted, the reaction partner, and the chemical structures, properties, and amounts of the products formed. According to photooxidation chamber experiments [*Odum et al*, 1996; *Hoffmann et al*. 1997, 1998; *Christoffersen et al.*, 1998; *Griffin et al.*, 1999a] the aerosol yield from oxidation reactions of α-pinene is about half that from β-pinene, which, in turn, is ~75% of that of limonene. Note, however, that the yields of these compounds are highly dependent on the amount of organic aerosol present in the experiment to act as an absorptive medium. It has been also shown that most of the mass of the oxidized monoterpenes remains in the gas phase [*Hoffmann et al*, 1998], so that monoterpenes also have an impact on gas phase chemistry, as taken into account in this study. It was further found that reactions of β-pinene with O_3 and nitrate radicals (NO_3), which are of major importance during night, lead to higher aerosol yields than reactions with hydroxyl radicals [*Hoffmann et al*, 1997], which dominate gas phase hydrocarbon oxidation during the day.

Odum et al [1996] and *Hoffmann et al* [1997] developed a simple model of gas-to-particle partitioning to describe the formation of organic particulate matter from biogenic hydrocarbon oxidation in chamber experiments. Very recently, *Griffin et al* [1999b] performed an exploratory study of the present-day SOA formation from the oxidation of BVOC, using the oxidant distributions derived from a three-dimensional (3-D) chemical transport model, and estimated a global production of SOA of 13-24 Tg/y SOA.

In this work, we present results from the first global 3-D modeling study of the evolution of the production and temporal and spatial distribution of SOA from the oxidation of biogenic hydrocarbons since the preindustrial times.

2. Model

2.1. Transport Model and Gas Phase Chemistry

We use the well-documented global three-dimensional chemistry-transport model, Model of the Global Universal Tracer Transport in the Atmosphere (MOGUNTIA) [*Crutzen and Zimmermann*, 1991; *Kanakidou and Crutzen*, 1993, 1999; *Dentener and Crutzen*, 1993], to simulate the spatial and temporal distribution of tropospheric oxidants, adopting the most recent kinetic data [*Hoffmann et al*, 1997; *Atkinson*, 1997; *Aschmann et al.*, 1998; *Hatakeyama et al*, 1991] and global emission inventories [*Guenther et al*, 1995] for biogenic organic compounds. The oxidant levels are calculated with a chemical scheme which describes carbon monoxide (CO), methane (CH_4), and nonmethane C_2-C_5 hydrocarbon (NMHC) oxidation, including isoprene [*Poisson et al*, 2000; *Kanakidou and Crutzen*, 1999]. This mechanism was expanded with a simplified gas phase pinene oxidation scheme, including O_3, OH, and NO_3 as oxidants. Ozonolysis of pinenes is also taken into account as a source of OH with yields of 0.76 and 0.33 from α- and β-pinene, respectively [*Atkinson*, 1997]. However, according to our calculations the overall global effect of pinenes on OH radicals is small. Enhanced conversion of nitric oxide (NO) to nitrogen dioxide (NO_2) by the additional generation of peroxy radicals (followed by photolysis of NO_2) acts as a source of ozone. Formation of CO, formaldehyde, and acetone from reactions of pinenes with OH and O_3 is taken into account on the basis of reported yields [*Atkinson*, 1997; *Aschmann et al.*, 1998; *Hatakeyama et al*, 1991]. Thus, in our model on a molar basis, reaction of α-pinene with OH yields 11% acetone and 97% CO; reaction of β-pinene with OH produces 8.5% acetone, 43% CO, and 54% formaldehyde. Ozonolysis of α-pinene is forming 22% formaldehyde and 9% CO whereas ozonolysis of β-pinene is producing 76% formaldehyde and 9% CO.

The model has been initialized by SOA, OA, and α- and β-pinene mixing ratios of 0.01 pptv. All other compounds have been set to December mixing ratios derived by previous simulations which reached equilibrium [*Kanakidou and Crutzen*, 1999]. The model is being run for a year with a spin-up time of a month.

2.2. Aerosol Formation and Removal

The aerosol/gas-phase-partitioning equilibrium model described by *Hoffmann et al* [1997] has been adopted for this study to describe SOA formation by biogenic volatile organic compounds. Because of substantial lack of data on the aerosol-forming efficiencies of the reactions with OH and NO_3, in this study we will only consider SOA formation by reactions of O_3 with α- and β-pinene, thus probably underestimating SOA production. Few mixed system kinetic data exist from the so-called "photochemistry" experiments, in which probably all three oxidants were present [*Hoffmann et al*, 1997; *Griffin et al*. 1999a]. While these studies indicate a potentially important contribution to SOA formation of other oxidants than ozone, it is unfortunately difficult to adopt these data for our study. Their use requires that the levels and the relative importance of the oxidants in the chamber are representative of the real atmosphere and close to the model calculations. Otherwise, the model results might favor oxidation of biogenic volatile organic compounds (BVOC) and thus SOA formation by the oxidant which is anomalously abundant in the chamber, providing, thus, erroneous estimate of SOA formation. Note, however, that the reactions of pinenes with OH and NO_3 radicals are taken into account as a sink for pinenes, affecting gas phase chemistry. Reactions of biogenics with ozone appear, however, to be the main source of SOA [*Hoffmann et al*, 1997]. This is in line with recent observations of sharp SOA rises in forested areas together with rapid increases in ozone [*Kavouras et al*, 1998, 1999; *Bonsang et al*, 1999]. SOA formation by ozonolysis is represented by a two-product formulation as shown below (equation (1)). These products are expected to be multifunctional compounds like diacids [*Saxena and Hildemann*, 1996; *Hoffmann et al*, 1998; *Kavouras et al.*, 1998]. The buildup of secondary organic aerosol mass is a function of the total equilibrium amount of OA that is available to absorb [*Odum et al.*, 1996; *Hoffmann et al*, 1997] and is calculated using an iterative solution as given by *Hoffmann et al* [1997]:

$$\begin{aligned} M_{o,t} = M_{o,n} = M_{o,t} & \\ & + M_{o,n-1} (\Delta\alpha \mathrm{PIN}_{O3} \{ [a_{11} K_{o11} / (1 + K_{o11} M_{o,n-1})] \\ & + [a_{12} K_{o12} / (1 + K_{o12} M_{o,n-1})] \} \\ & + \Delta\beta \mathrm{PIN}_{O3} \{ [a_{21} K_{o21} / (1 + K_{o21} M_{o,n-1})] \\ & + [a_{22} K_{o22} / (1 + K_{o22} M_{o,n-1})] \}) \qquad (1) \end{aligned}$$

where $M_{o,t}$ is the aerosol mass at time t, $M_{o,t}$ is the initial aerosol mass of SOA, $n=1\rightarrow\infty$, $\Delta\alpha \mathrm{PIN}_{O3}$ and $\Delta\beta \mathrm{PIN}_{O3}$ are the amounts of α- and β-pinenes that reacted with O_3 in time step Δt, and a_{ij} and K_{oij} are the kinetic aerosol production parameters shown in Table 1. In particular, a_{ij} is the mass-based stoichiometric yield of the condensable compound j produced by the ozonolysis of BVOC i, and K_{oij} is the gas-particle equilibrium coefficient used to describe the partitioning of this product between the aerosol and the gas phase [*Odum et al.*, 1996].

Two different sets of kinetic aerosol production parameters (Table 1) have been used to evaluate the sensitivity of the results to these parameters. Furthermore, the term M_o may contain not only the mass of SOA but also the primary organic aerosol (OA). Two different simulations have, therefore, been performed with different assumptions regarding M_o used in (1). In the first case, partitioning of the organic matter takes place only on SOA generated by BVOC oxidation, which is closest to the conditions in the chamber experiments used to derive (1). For the second simulation we also allowed deposition of condensable compounds on all preexisting OA (primary OA plus the SOA resulting from anthropogenic organic compound oxidation), assuming that all organics formed during BVOC oxidation have the same affinity for condensation on OA as on SOA. We refer hereinafter to the latter scenario as the base case, and the thus-calculated distributions of SOA are presented in section 3.1. It should be noted that deposition onto other aerosol types and absorption into aqueous aerosols via Henry's law are not taken into account in this study.

Table 1. Stoichiometric and Partitioning Coefficients to Parameterize SOA Formation From Ozonolysis

	Griffin et al [1999a]	*Hoffmann et al* [1997]
	α-pinene	
a_{11}	0.125	0.12
K_{o11}	0.088	0.2
a_{12}	0.102	0.19
K_{o12}	0.0788	0.005
	β-pinene	
a_{21}	0.026	0.35
K_{o21}	0.195	0.11
a_{22}	0.485	
K_{o22}	0.003	

A mean molecular weight of 170 for SOA has been used

For simplification, in all simulations presented in section 3, the mixture of BVOC (i.e., monoterpenes include α- and β-pinene, limonene, myrcene, sabinene, Δ-carene, ocimene, terpinolene, and terpinene, and the aerosol-producing other reactive volatile organic compounds (ORVOC) include terpenoid alcohols, *n*-carbonyls, aromatics, sesquiterpenes, terpenoid ketones, and higher olefins) has been represented by equal masses of α- and β-pinene, since these compounds are the major contributors (by approximately 35% and 25%, respectively) to the global monoterpene emissions [*Guenther et al.*, 1995]. The use of a global percentage is justified by the lack of data on the emission distribution of the individual BVOC on a global scale. Recently, *Griffin et al.* [1999b] have performed a compilation of the few existing, sporadic data of the individual compounds of the emissions of BVOC on global scale, which is used in the present study to calculate efficient reaction rates for α- and β-pinene (see Appendix). Moreover, α- and β-pinene were chosen to represent all BVOC that can produce SOA because they have different characteristic behaviors regarding aerosol formation. The α-pinene belongs to the category of BVOC which have a double bond internal to the ring structure, whereas β–pinene has a terminal $=CH_2$ group outside the cyclic monoterpene structure. The oxidation of α–pinene will thus lead to cleavage of the six-member ring but will conserve the number of carbon atoms, and the product will have two functional groups. On the other hand, β–pinene oxidation will lead to products with less carbon atoms and one functional group [*Griffin et al*, 1999a]. The other compounds which can be important contributors to SOA formation, as will be discussed in section 3.3, are also represented by these two pinenes. This approximation may lead to an underestimate of the global SOA formation by our model. However, in order to achieve a representative reactivity for the thus-lumped terpenes, we adopted effective rate constants to account for the BVOC other than α- and β-pinene. These effective rate constants are derived in the Appendix for two different cases of lumped BVOC.

The impact of wet and dry deposition processes on organic aerosols was taken into account. Assuming that the bulk of SOA resides in the accumulation range, a mean deposition velocity of 0.1 cm/s has been adopted [*Liousse et al*, 1996]. The wet removal of SOA is parameterized following *Junge and Gustafson* [1957] and *Langner and Rodhe* [1991], by whom the removal rate L_p is taken to be proportional to the average rate of formation of precipitation R ($g/m^3/s$):

$$L_p = \varepsilon R_o g / L$$

where ε is a dimensionless parameter describing the efficiency of the uptake of the aerosol into the precipitation elements, R_o is the precipitation rate at the Earth's surface (in $g/m^2/s$), g is the fraction of precipitation released in a given height interval and calculated from the zonal mean data on the release of latent heat (in m^{-1}) [*Newell et al.*, 1974], and L is the liquid water content (g/m^3) of precipitating clouds.

Unfortunately, little is known about the hygroscopic properties of SOA and how this affects its removal by precipitation. In this work we assumed a rather efficient removal, comparable to that of the highly soluble sulfate aerosol [*Langner and Rodhe*, 1991]. We implicitly assume that because of the presence of soluble organic acids of the SOA and because

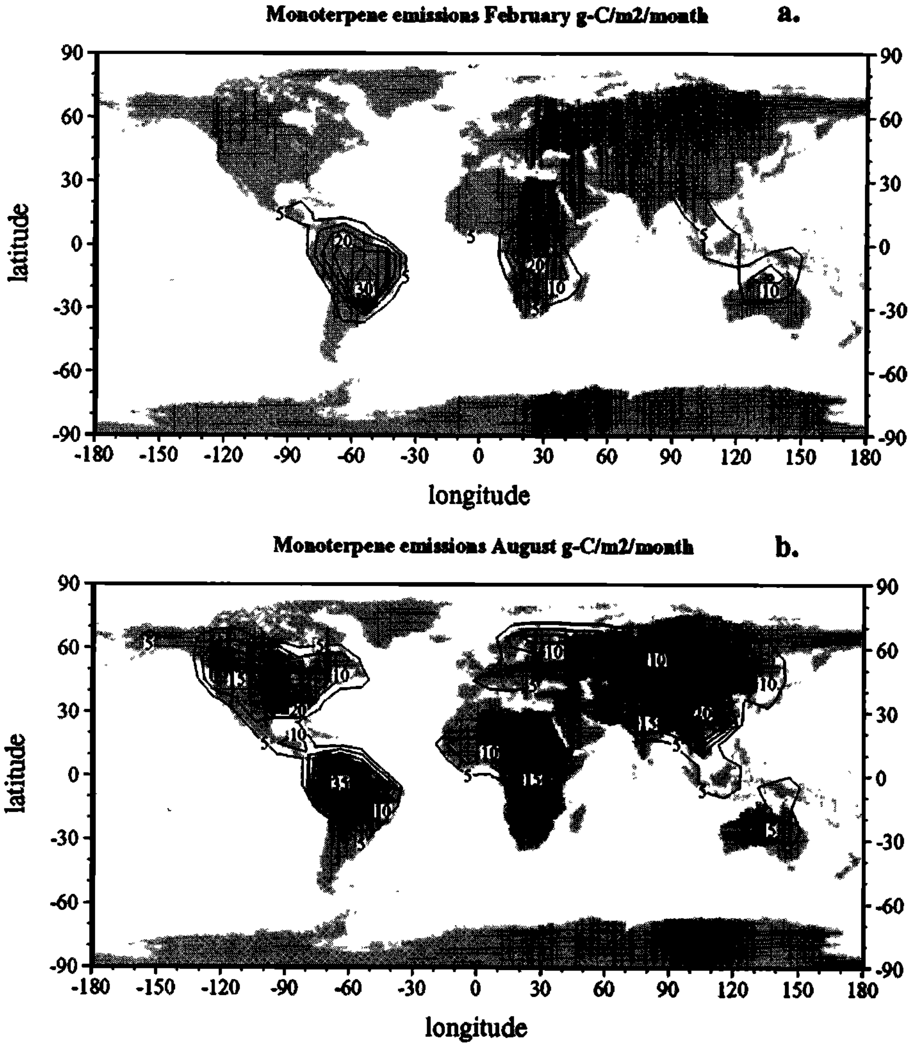


Figure 1. Monoterpene emission distributions adopted from *Guenther et al* [1995] (a) in February and (b) in August in g-C/m^2/month.

of mixing with highly soluble sulfates and nitrates, the aerosol behaves hygroscopically. Thus the calculated SOA removal by precipitation might be overestimated, but this uncertainty does not affect our assessment of the change of preindustrial versus industrial SOA production rates. Unfortunately, in our opinion, at this moment a more sophisticated removal parameterization is not supported by data. Obviously, this assumption has to be refined in future studies when additional knowledge on the chemical composition and the physicochemical behavior of the SOA will be gained.

2.3. Present-Day and Preindustrial Boundary Conditions

Global emissions of volatile organic compounds by vegetation, which are estimated by *Guenther et al* [1995] to be equal to 1150 Tg-C/yr, are highly uncertain but, in any case, much higher than the anthropogenic hydrocarbon emissions of 120 Tg-C/yr [*Olivier et al*, 1996]. Of the former, monoterpenes contribute with 127 Tg-C/yr (Figure 1), and other reactive volatile organic compounds excluding isoprene (ORVOC) contribute with 260 Tg-C/yr. Aerosol is produced from the oxidation products of monoterpenes (127 Tg-C/yr) and of the most reactive and aerosol-producing part of ORVOC (32% or 83 Tg-C/yr) [*Griffin et al*, 1999b]. These emissions (210 Tg-C/yr) have been adopted in the model to calculate SOA formation. As explained in section 2.2, the BVOC are assumed to be emitted as a mixture of α- and β-pinene only. Emissions of organic aerosols derived from fossil fuel (28.5 Tg/yr) and from tropical biomass burning (44.6 Tg/yr) were treated in the model as direct emissions [*Liousse et al.*, 1996] and constitute the only categories of anthropogenic aerosol emissions in the model. The fossil fuel and biomass-burning emission distributions by *Crutzen and Zimmermann* [1991] have been used to distribute these primary particulate emissions.

To evaluate the impact of human activities on SOA production by the oxidation of BVOC we neglected in our preindustrial simulation the anthropogenic emissions of CH_4, NO_x, CO, nonmethane hydrocarbons, acetone, and fossil-fuel-derived OA. We also reduced the biomass-burning emissions of all trace compounds and OA to 10% of their present value and further adopted the preindustrial soil emission estimates of NO_x [*Yienger and Levy*, 1995]. The emissions of monoterpenes were kept constant between preindustrial times and the present, although the Earth's forested areas may have decreased between 1850 and 1980 by up to 20% [*Enquete Commission of the 11th German Bundestag*, 1990; *International*

Panel on Climate Change (*IPCC*), 1992; *Houghton et al*. 1983]. *Houghton et al* [1983] estimated that <20% reduction took place in the northern midlatitude to high-latitude forested areas between 1700 and 1980. Thus the assumption of constant biogenic emissions between the preindustrial period and the present could introduce an underestimate of roughly 20% in hydrocarbon emissions from forests. Our calculations will, however, show that the impact of deforestation is of minor importance compared to the SOA enhancement by other human activity.

3. Results

In section 3.1 we present results for our base case simulation (see section 2.2) and compare them to measurements. To show the combined influence of seasonally varying oxidant concentrations and natural hydrocarbon emissions, we contrast the results obtained for February (winter Northern Hemisphere (NH)) and August (summer NH). In section 3.2 we evaluate the sensitivities to the various assumptions made in this study; these sensitivities and combinations of them are summarized in Table 2. A comparison of our present-day results with other works is made in section 3.3 where global burdens are also given. Finally, the difference between preindustrial and present-day simulations is presented and discussed in section 3.4.

3.1. Calculated Distributions

The calculated SOA distribution for the industrial period (simulation a in Table 2 using aerosol-forming parameters from *Hoffmann et al* [1997], Figure 2a for February, and Figure 3a for August) shows maximum values over tropical forested areas, the location of maximum biogenic hydrocarbon emissions. There, the calculated values, exceeding 20 $\mu g/m^3$ in August and 10 $\mu g/m^3$ in February, compare with the 0.5-20 $\mu g/m^3$ of total organic carbon measured in the tropics (see references given by *Liousse et al.* [1996]). During summer, terpene emissions maximize over forested midlatitude areas in the Northern Hemisphere (Figure 1b), and this is reflected in the SOA distributions. High concentrations up to 15 $\mu g/m^3$ of SOA are thus calculated in these regions for August where total organic carbon has been observed to range from 1 to 16 $\mu g/m^3$ (*Bonsang et al.* [1999] and references given by *Liousse et al.* [1996]). Much lower concentrations of SOA of the order of 1 $\mu g/m^3$ are calculated for February due to lower BVOC emissions and oxidant levels during the NH winter. Note that boreal forest fires have been neglected as a source for OA in this study since not enough data are available regarding the intensity and the distribution of this source.

Table 3 shows comparisons of model-calculated SOA and OA concentration to the few available surface observations at various forested and rural locations around the globe for the "base" simulation (simulation a in Table 2). Model organic aerosols (OA), shown in Figures 2b and 3b, represent here the sum of biogenic SOA plus the OA derived from the industrial and biomass-burning sources given by *Liousse et al* [1996]. The model comparison with observations focuses on continental areas, from which the biogenic precursor gases of SOA largely originate. As shown in Table 3, the model is able to reproduce within a factor of 2-4 the observed OA in most of the continental rural and forested areas. However, it underestimates the OA observations at Pertouli, a forested mountain site in Greece at 1.3 km elevation. This location, depending on the time of the day and weather conditions, can be affected by subgrid pollution which cannot be resolved in our model.

Despite the coarse grid of our model (1000 x 1000 km) the model results are encouraging. The vertical resolution of MOGUNTIA (first layer roughly 450 m) implies a perfect mixing within this layer; while during the day under turbulent conditions this is a reasonable approach, during the night, mixing may be somewhat overestimated. Likewise, the model is probably more representative for summer situations when all emissions are high than for winter, associated with shallow and stable boundary layers. Improvement of the model will require higher resolution and explicit consideration of SOA production from anthropogenic organic compound oxidation. The contribution of the SOA from BVOC oxidation to the total OA is shown by comparison of Figures 2a and 2b for February and Figures 3a and 3b for August. It can be seen that regionally, the SOA from the ozonolysis of BVOC contributes more than 40-60% to the total OA in the tropics and most continental regions at northern midlatitudes.

3.2. Uncertainties and Sensitivities

Because of the uncertainties in the emissions of the BVOC and because not all the compounds forming SOA have been identified, the sensitivity of the model results to the global annual emissions of the precursors of SOA has to be investigated. For this purpose, the global production of SOA from biogenic compound oxidation has been calculated for the industrial period for two different amounts of emissions. The first set of simulations (simulations a, b, e, and f in Table 3) has been performed by taking into account the 127 Tg-C/yr of annual emissions of monoterpenes and 83 Tg-C/yr of the very efficient SOA-producing ORVOC. In the second set (simulations c, d, g, and h) only the 127 Tg-C/yr of annual global emissions of monoterpenes are considered [*Guenther et al*, 1995]. As expected, in the latter case, lower emissions of organic SOA precursors and lower chemical reactivity of the hydrocarbon mixture lead to much lower calculated SOA formation, 0.08 versus 45.8 Tg/yr, when only considering self-nucleation and deposition on SOA and kinetic aerosol data by *Hoffmann et al.* [1997] (Table 2, simulations d and b, respectively). Remarkably, when partitioning on all available OA is allowed (simulations a and c), a strong increase of aerosol yield is calculated compared with the previous simulations for, especially, the monoterpene simulation (simulation c compared with simulation d). Another factor of uncertainty comes from the temperature dependence of the aerosol yield. A number of factors affecting aerosol formation depend on temperature, such as the partitioning coefficients of the condensable compounds, which are negatively correlated with temperature [*Odum et al.*, 1996; *Hoffmann et al.*, 1997], the emissions of BVOC, which increase by approximately 9% K^{-1} [*Guenther et al*, 1995], and the rates of oxidation reactions of BVOC, which also increase with temperature. The temperature dependence of reaction rates and emissions is taken into account in our simulations, except for the diurnal temperature variations. Because of lack of appropriate information, no temperature dependence of the partitioning coefficients has been considered in this study, although several of the chamber experiments, from which the parameters for (1) were derived, were conducted at temperatures higher than 40°C [*Hoffmann et al*, 1997]. This overestimation of temperatures may imply a substantial underestimation of the efficiency of SOA formation.

The adopted aerosol-forming parameters could be a large source of uncertainty in our calculations. Thus simulations have been also performed by adopting the aerosol-forming parameters of α– and β–pinene, which are used to represent the mixture of

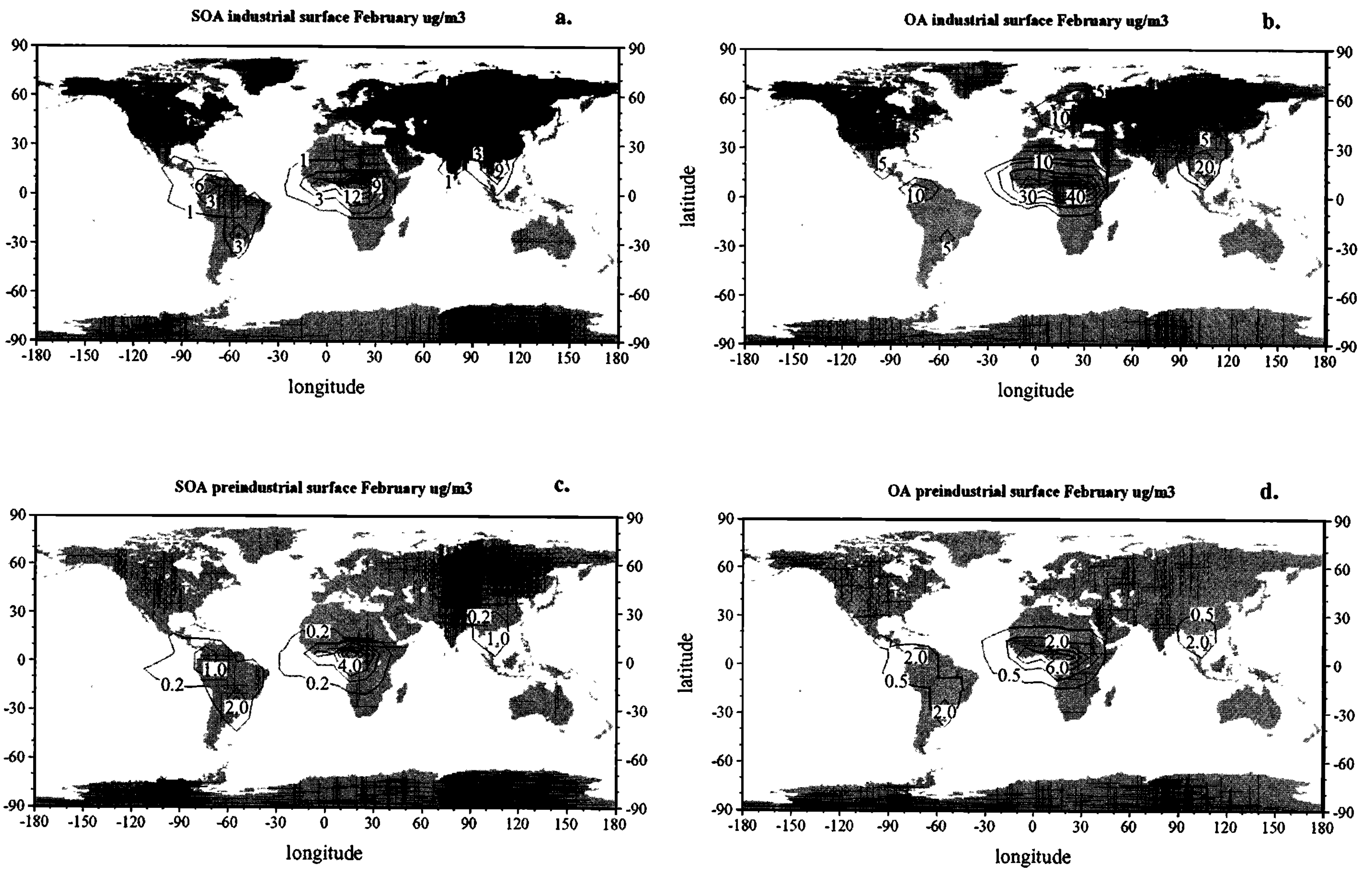

SOA industrial surface February ug/m3
a.
OA industrial surface February ug/m3
b.
SOA preindustrial surface February ug/m3
c.
OA preindustrial surface February ug/m3
d.
latitude
longitude
90
60
30
0
-30
-60
-90
-180
-150
-120
-90
-60
-30
0
30
60
90
120
150
180
1
3
6
9
12
5
10
20
30
40
0.2
1.0
2.0
4.0
0.5
6.0

Figure 2. Calculated surface distributions for February in μg/m^3: (a) present-day SOA, (b) present-day OA (sum of SOA and OA derived from industrial and biomass-burning sources), (c) preindustrial SOA, and (d) preindustrial OA.

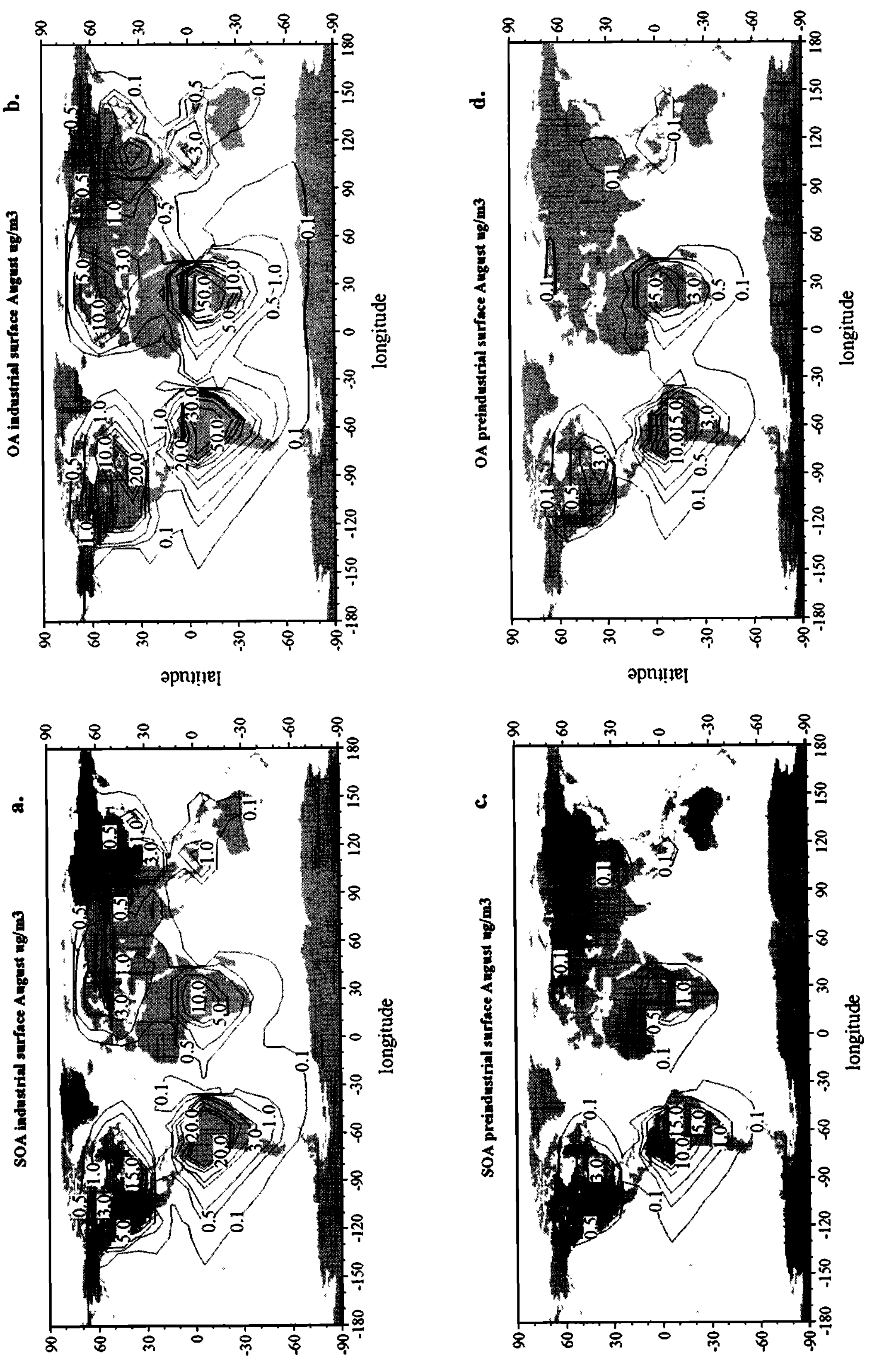

a.
SOA industrial surface August ug/m3
b.
OA industrial surface August ug/m3
c.
SOA preindustrial surface August ug/m3
d.
OA preindustrial surface August ug/m3
latitude
longitude
90
60
30
0
-30
-60
-90
-180
-150
-120
-90
-60
-30
0
30
60
90
120
150
180

Figure 3. Calculated surface distributions for August in µg/m^3: (a) present-day SOA, (b) present-day OA, (c) preindustrial SOA, and (d) preindustrial OA.

Table 2. Sensitivity of SOA Formation to Emissions and the Chemical Reactivity of the Mixture of Biogenic Hydrocarbons for the Industrial and Preindustrial Simulations

Simulation	Simulation Description	*Griffin et al* [1999a]		*Hoffmann et al* [1997]	
		Burden	Production	Burden	Production
			Industrial Period		
a	base BVOC, partitioning on OA	1.2	61.4	1.6	78.9
b	BVOC, partitioning on SOA	0.6	29.4	0.9	45.8
c	monoterpenes, partitioning on OA	0.2	8.1	0.2	10.5
d	monoterpenes, partitioning on SOA	<0.1	0.7	<0.1	0.08
			Preindustrial Period		
e	base BVOC, partitioning on OA	0.4	16.5	0.6	27.9
f	BVOC, partitioning on SOA	0.3	11.9	0.5	22.1
g	monoterpenes, partitioning on OA	<0.1	1.3	<0.1	1.6
h	monoterpenes, partitioning on SOA	<0.1	0.4	<<	<0.01

Simulation a, base simulation as presented in sections 2.3 and 3.1, considering SOA formation by BVOC (monoterpenes and ORVOC) and assuming condensation on all OA. Thus both pure SOA particles from BVOC oxidation and mass attached on preexisting OA particles are considered. The α-pinene and β-pinene reactivities versus O_3, OH, and NO_3 radicals are increased by 20, 2, and 3 times, respectively, to account for the more reactive aerosol-producing BVOC (see Appendix). Simulation b, same as simulation a, but assuming only self-nucleation and condensation on SOA. Simulation c, considering SOA formation only from monoterpenes and assuming condensation on all OA. Thus α-pinene and β-pinene reactivities versus O_3, OH, and NO_3 radicals are increased by 5.2, 1.6, and 2.3 times, respectively, to account for the more reactive monoterpenes as explained in the Appendix. Simulation d, as for simulation c but assuming only self-nucleation and condensation on SOA. Simulation e, same as simulation a, but for preindustrial times, emissions as presented in 2.3. Simulation f, same as simulation e, but assuming only self-nucleation and condensation on SOA. Simulation g, same as simulation c, but for preindustrial times. Simulation h, same as simulation g, but assuming only self-nucleation and condensation on SOA. Burden, annual mean SOA burden in Tg up to 100 hPa, production, annual chemical production of SOA from BVOC oxidation in Tg/yr.

biogenics, suggested by *Griffin et al.* [1999a] and shown in Table 1. The results are also presented in Table 2. The use of these parameters reduced the SOA production for the case of self-nucleation and partitioning on SOA (simulation b) of 45.8 (using data from *Hoffmann et al* [1997] (hereinafter referred to as H97)) to 29.4 Tg/yr (using data from *Griffin et al* [1999a] (hereinafter referred to as G99)). Allowing for deposition also on OA (simulation a), the corresponding numbers increase to 78.9 and to 61.4 Tg/yr respectively.

3.3. Global Budget

Globally, we calculate a production range of 61-79 Tg-SOA/yr from the ozonolysis of BVOC, including ORVOC, allowing for partitioning of condensable compounds onto other OA from primary sources and anthropogenic VOC oxidation (simulation a). About one third of these values calculated if only self-nucleation and deposition on SOA are considered (simulation b). This range in global aerosol production results from the use of two different sets of parameters of aerosol production from ozonolysis of α- and β-pinene [*Hoffman et al*, 1997; *Griffin et al*, 1999a] (Table 1). The global burden of SOA up to 100 hPa for the base case (simulation a) is calculated to be 1.2-1.6 Tg. Comparison of these amounts with the 0.09 Tg OA from fossil fuel [*Cooke et al.*, 1999], the 2.9-3.2 Tg OA from all sources (this work), and the 1.14-2.26 Tg $SO_4^=$ [*Langner and Rodhe*, 1991; *Chin et al.*, 1996; *Feichter et al*, 1996, 1997] shows the great potential importance of SOA from BVOC oxidation in the atmosphere. Our results raise fundamental questions on what the amount and the composition of preindustrial aerosol were, and how this affected cloud and precipitation formation.

Except for the neglect of aerosol formation due to reactions of the BVOC with OH and NO_3 and possible temperature effects, the numbers given above are upper limits to the possible SOA production and are much dominated by the ozonolysis of the highly reactive ORVOC. If the latter components are ignored, we only calculate a SOA production of 10.5 Tg/yr even if we take into account deposition of products on all available

Table 3. Comparison of Model Results (Simulation a, *Hoffmann et al.* [1997] in Table 2) to Observations Over Continental Areas

Location	Time Period	Observed	Calculated		References
		OA	SOA	OA	
		Forest			
40.4°N, 8°W (Tabua)	July-Aug	1.34-7.16	0.71	1.91	1
39.5°N, 21.5°E (Pertouli)	July-Aug	3.93-7.05	0.66	1.59	1
		Rural			
71.1°N, 156.2°W (Barrow)	Jan.-April	1.31	0.02	0.36	2
40°N, 79°W (Allegheny)	Aug	2.15	5.31	7.63	2
44°N, 1°W (Lands)	annual	1.50	0.26	1.45	2
6.21°N, 5.03°W (Lamto)	Nov.-March	2.00-20.00	6.72	24.82	3
6.21°N, 5.03°W (Lamto)	annual	5.60	3.24	11.08	3
2°S, 77.3°W (Ecuador)	annual	0.51	0.48	0.61	4
10°S, 76°W (Peru)	March-April	0.16	0.04	0.04	4
48.42°N, 2.8°E (Gif sur Yvette)	annual	5.75	1	6.26	3

References 1, *Bonsang et al* [1999], 2, references given by *Cooke et al* [1999], 3, *Cachier et al* [1996], 4, references given by *Liousse et al* [1996]. Units are μg/m^3

OA. This strongly demonstrates the great uncertainty in current estimates and the importance of identifying the chemical composition, reactivity, particle formation properties, and capability of deposition on OA and maybe other aerosol, of the many oxidation products of the organic compounds that are emitted by the world's vegetation.

While this study was in progress, we obtained a preprint of the paper by *Griffin et al* [1999b] in which they calculated the present-day SOA production by BVOC, off-line, using the results of a global model similar to MOGUNTIA, the Intermediate Model of the Global and Annual Evolution of Species (IMAGES) model [*Muller and Brasseur*, 1995]. *Griffin et al* [1999b] used fixed oxidant fields derived by the IMAGES model, which is not the case in our study, where oxidants are calculated interactively. For the SOA calculations which were performed off-line, they used vertical layers based on the chemical reactivities of the studied species. Thus vertical cells were defined by increasing values of the scale heights for the compounds of interest (see details given by *Griffin et al* [1999b]). In our study, this differentiation is not considered since the reaction products of BVOC live probably long enough to be mixed into the model domain, and the photochemically produced SOA is transported consistently with the MOGUNTIA model transport pattern. *Griffin et al.* [1999b] performed their calculations based on SOA releases from 13 BVOC, including the highly reactive ORVOC. The aerosol-forming parameters they used in their study were based on their own photochemistry experiments under sunlight conditions. Therefore in that work they also considered the contribution of OH reactions to SOA formation, whereas we have ignored it for the time being. However, as argued in section 3.2, there are considerable uncertainties related to the application of such data to the real atmosphere. In our study the calculations are based on the ozonolysis parameters which were derived for α- and β-pinenes [*Hoffmann et al*, 1997; *Griffin et al*, 1999a], with reactivities adjusted for the other BVOC, including or excluding the ORVOC. *Griffin et al* [1999b] obtain a global production of 13-24 Tg/yr of SOA, allowing only for self-nucleation and partitioning on SOA (no deposition on other OA). These numbers can be compared with our simulations (simulation b) and agree reasonably well with our 29-46 Tg/yr when considering the above-explained different approaches used in

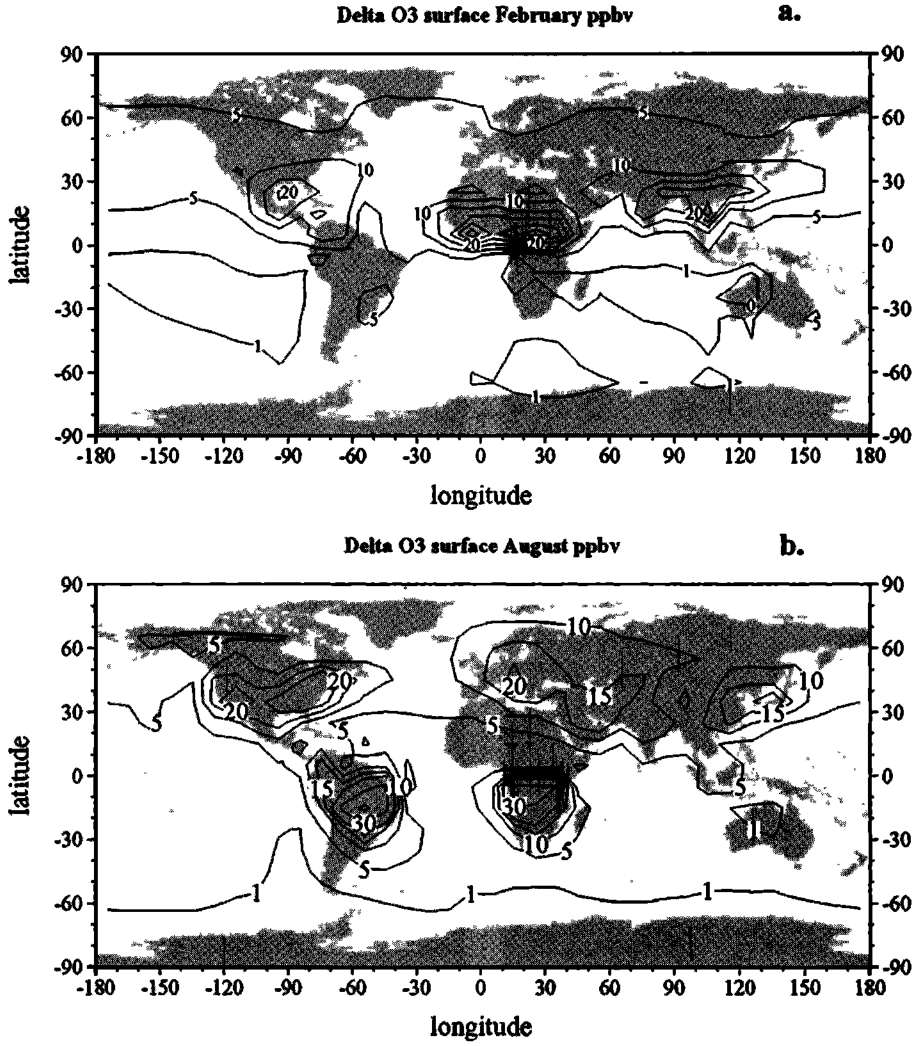


Figure 4. Increase of surface O_3 mixing ratios (in ppbv) between present day and preindustrial period (a) in February and (b) in August.

the two studies. The two studies reveal the very large sensitivity to critical properties, such as hydrocarbon mixture, whether or not we allow deposition on OA and inclusion or exclusion of oxidants having the potential to initiate SOA formation.

3.4. Preindustrial Versus Present-Day Simulations

Although originating from natural sources, SOA concentrations for the industrial period are much higher than the computed levels for the preindustrial era (Figures 2c and 3c for February and August, respectively). It is striking that according to our calculations only relatively little organic aerosol formation occurred over the NH forested areas before human activity injected nitrogen oxides and hydrocarbons in the troposphere, leading to an increase in oxidant levels, in particular O_3 [*Crutzen and Zimmermann*, 1991; *Volz and Kley*, 1988] (see Figure 4). The calculated relative increases in SOA and OA concentrations between the preindustrial era and the present maximize in the NH as can seen in Figure 2 for February and Figure 3 for August. Human activity in these areas also injected primary organic particles on which condensable organics can attach. Maximum absolute increases are calculated for the tropics where anthropogenic biomass burning enhanced both O_3 and primary OA concentrations. Comparison between Figures 2c and 2d for February and Figures 3c and 3d for August shows that SOA constituted the major fraction of OA (which includes SOA) in preindustrial times, although O_3 concentrations were substantially lower, because only little OA was injected into the atmosphere by primary sources.

According to our calculations for deposition on OA (simulation e, Table 3) only 16.5-27.9 Tg/yr of SOA were produced from BVOC oxidation in the preindustrial period These numbers are almost 3-4 times less than the 61.4-78.9 Tg/yr calculated for SOA production from BVOC oxidation in the industrial period (simulation a). Thus ~75% of the present SOA formation from naturally emitted biogenic compounds (Table 3) may be caused by human activity.

Although nowadays sulfate aerosols are mostly believed to be major constituents of CCN on a global scale, the water-soluble organic aerosol fraction could be important in CCN formation over and downwind of forested areas [*Novakov and Penner*, 1993; *Novakov and Corrigan*, 1996; *Andreae and Crutzen*, 1997: *Talbot et al*, 1988, 1990: *Kavouras et al*, 1998]. In the preindustrial atmosphere, SOA might have been the major component of CN and, possibly, CCN over continental areas. Our calculations, nevertheless, indicate that SOA production may have increased threefold to fourfold since preindustrial times because of reactions of biogenic compounds with anthropogenically produced O_3 and through deposition on organic particles of anthropogenic origin. Note that deposition on other preexisting particles like sulfate or soil-derived aerosols, which could further enhance SOA formation, has been neglected in this study.

4. Discussion and Conclusions

About 75% of contemporary secondary organic aerosol formation from ozonolysis of naturally emitted biogenic compounds, which is calculated to be in the range of 61-79 Tg/yr, may be due to human activities. This organic aerosol source has increased threefold to fourfold since preindustrial times. This enhancement of SOA might have induced changes in cloud properties, the hydrological cycle, and cloud albedo and, thereby, in climate. SOA contributes presently ~50% to the total organic aerosol sources whereas, during the preindustrial period, this source, although substantially weaker than nowadays, may have contributed ~80% to the global budget of suspended organic matter and may thus have constituted the major source of cloud condensation nuclei over the continents. This raises fundamental questions as to the amount and composition of preindustrial aerosol and how this affected the formation of clouds and rain. Because of the importance of organic aerosols for atmospheric chemistry and climate, it is very important to better quantify the natural emissions and particle-producing properties of all relevant organic compounds. More laboratory and field studies are needed to understand the chemical reactions of biogenic organics and products, including condensables, and the subsequent formation of SOA. Future growth of human activities in the tropics and subtropics, leading to enhanced emissions of NO_x and primary organic matter from fossil fuel and biomass burning can have a major impact on organic particulate concentrations, tropospheric chemistry, and various climate factors.

Appendix

BVOC other than α- and β-pinene can react from 2 to 150 times faster with ozone than the pinenes [*Atkinson*, 1997] while their reactivity versus OH and NO_3 radicals can be ~1.2 to 10 times faster for reactions with OH and from 2 to 50 times faster for reactions with NO_3 radicals [*Atkinson*, 1997]. To correct for this effect and represent all monoterpenes (Table 3, simulations c, d, g, and h) by a mixture of α- and β-pinene, we increased the chemical reactivity of α- and β- pinene with respect to O_3, OH, and NO_3 by factors of 5.2, 1.6, and 2.3, respectively, based on the monoterpene emission composition [*Griffin et al*, 1999b] and reaction coefficients [*Atkinson*, 1997] without changing the aerosol production parameters. These factors F have been calculated as the ratio of the mean reactivity of the monoterpenes, weighted by their contribution to the total monoterpenes emissions, to the reactivity of a mixture containing 50% α-pinene and 50% β-pinene:

$$F = \Sigma\, k_{xi}\, a_i \,/\, (0.5\, k_{x\alpha\mathrm{PIN}} + 0.5\, k_{x\beta\mathrm{PIN}}),$$

where k_{xi} is the rate of the reaction of the monoterpene i with the oxidant X, a_i is the contribution of the monoterpene i to the total monoterpene emissions, and $k_{x\alpha\mathrm{PIN}}$ and $k_{x\beta\mathrm{PIN}}$ are the rates of the reactions of α– and β–pinene with the oxidant X).

Similarly, for the simulation also including the other highly reactive organic compounds (ORVOC) like sesquiterpenes, terpenoid alcohols, and ketones [*Griffin et al*, 1999b] (Table 3, simulations a, b, e, and f), we increased the reactivity of both α- and β-pinene by factors of 20, 2, and 3 for reactions with O_3, OH, and NO_3, respectively. As it turned out, in the results presented in Table 3 and in section 3, the ORVOC are largely responsible for SOA formation. Note that these higher rate coefficients increase the oxidation rates of the α– and β–pinene surrogates for BVOC by all oxidants, but only the amounts of BVOC that reacted with O_3 ($\Delta\alpha\mathrm{PIN}_{O3}$ and $\Delta\beta\mathrm{PIN}_{O3}$) are used in formula (1).

Acknowledgments. M K. and K T acknowledge LSCE for its hospitality, IDRIS for computing facilities, and the Committee for Research of the University of Crete for support We also wish to thank the two anonymous reviewers for their constructive comments This research has been supported by the EU ENV4CT970483 SINDICATE program

References

Andreae, M O , and P J Crutzen, Atmospheric aerosols Biogeochemical sources and role in atmospheric chemistry, *Science*, *276*, 1052-1058, 1997

Artaxo, P , H Storms, F Bruynseels, R V Grieken, and W Maenhaut, Composition and sources of aerosols from the Amazon Basin, *J Geophys Res* , *93*, 1605-1615, 1988

Artaxo, P , W Maenhaut, H Storms, and R V Grieken, Aerosol characteristics and sources for the Amazon Basin during the wet season, *J Geophys Res* , *95*, 16,971-16,986, 1990

Aschmann, S M , A Reissel, R Atkinson, and J Arey, Products of the gas-phase reactions of the OH radical with a- and ß-pinene in the presence of NO, *J Geophys Res* , *103*, 25,553-25,561. 1998

Atkinson, R , Gas-phase tropospheric chemistry of volatile organic compounds, 1, Alkanes and alkenes, *J Phys Chem Ref Data*, *26*, 215-290, 1997

Bonsang, B , E Stephanou, C Pio, and P Seakins, Aerosols Formation from Biogenic Organic Carbon (AEROBIC), *Final Rep . Contract ENV4-CT95-0049*, R&D Programme Env and Clim , Eur Comm - DG XII, Brussels, June 1999

Cachier, H , C Liousse, M H Pertuisot, A Gaudichet, F Echalar, and J P Lacaux, African fire particulate emissions and atmospheric influence, in *Biomass Burning and Global Change*, edited by J S Levine, pp 428-440, MIT Press, Cambridge, Mass 1996

Chin, M , D J Jacob, G M Gardner, M S Foreman-Fowler, P A. Spiro, and D L Savoie, A global three-dimensional model of tropospheric sulfate, *J Geophys Res* . *101*, 18,667-18,690, 1996

Christoffersen, T S , et al , Cis-pinic acid, a possible precursor for organic aerosol formation from ozonolysis of a-pinene, *Atmos Environ* , *32*, 1657-1661, 1998

Cooke, W , C Liousse, H Cachier, and H Feichter, Construction of a one degree x one degree fossil fuel emission data set for carbonaceous aerosol and implementation and adiative impact in the ECHAM 4 model, *J Geophys Res* , *104*, 22,137-22,162, 1999

Crutzen, P J , and P H Zimmermann, The changing photochemistry of the troposphere, *Tellus Ser A*, *43*, 136-151, 1991

Cruz, C N , and S N Pandis, A study of the ability of pure secondary organic aerosol to act as cloud condensation nuclei, *Atmos Environ* , *31*, 2205-2214, 1997

Dentener, F J , and P J Crutzen, Reaction of N_2O_5 on tropospheric aerosols. Impact on the global distributions of NO_x, O_3, and OH, *J Geophys Res* , *98*, 7149-7163, 1993

Enquete Commission of the 11th German Bundestag, *Protecting the Tropical Forests*, edited by Deutscher Bundestag Referat Öffentlichkeitsarbeit, 968 pp., Enquete Commission of the 11th German Bundestag, Bonn, 1990

Feichter, J , E Kjellström, H Rodhe, F J Dentener, J Lelieveld, and G -J Roelofs, Simulation of the tropospheric sulfur cycle in a global climate model, *Atmos Environ* , *30*, 1693-1707, 1996

Feichter, J , U Lohmann, and I Schult, The atmospheric sulfur cycle in ECHAM-4 and its impact on the shortwave radiation, *Clim. Dyn* , *13*, 235-246, 1997

Griffin, R J., D R Cocker III, R C Flagan, and J H Seinfeld, Organic aerosol formation from the oxidation of biogenic hydrocarbons, *J Geophys Res.*, *104*, 3555-3567, 1999a

Griffin, R.J., D R Cocker III, J H Seinfeld, and D Dabdub, Estimate of global atmospheric organic aerosol from oxidation of biogenic hydrocarbons, *Geophys Res Lett* , *26*, 2721-2724, 1999b

Guenther, A , et al , A global model of natural volatile organic compound emissions, *J Geophys Res* , *100*, 8873-8892, 1995

Hatakeyama, S , K Izumi, T Fukuyama, H. Akimoto, and N Washida, Reactions of OH with *a*-pinene and ß-pinene in air Estimate of global CO production from atmospheric oxidation of terpenes, *J Geophys Res* , *96*, 947-958, 1991

Hoffmann, T , J R Odum, F Bowman, D Collins, D Klockow, R C Flagan, and J H Seinfeld, Formation of organic aerosols from the oxidation of biogenic hydrocarbons, *J Atmos Chem* , *26*, 189-222, 1997

Hoffmann, T., R Bandur, U Marggraf, and M Linscheid, Molecular composition of organic aerosols formed in the *a*-pinene/O_3 reaction Implications for new particle formation processes, *J Geophys Res* , *103*, 25,569-25,578, 1998

Houghton, R A , J E Hobbie, and J M Melillo, Changes in the carbon content of terrestrial biota and soils between 1860 and 1980 A net release of CO_2 to the atmosphere, *Ecol Monogr* , *53*(3), 235-262, 1983

Houweling, S , F J Dentener, and J Lelieveld, The impact of nonmethane hydrocarbons on tropospheric photochemistry, *J Geophys Res* , *103*, 10,673-10,696, 1998

International Panel on Climate Change (IPCC), *Climate Change 1992*, edited by J T Houghton, Cambridge Univ Press, New York, 1992

Junge, C E , and P E Gustafson, On the distribution of seasalt over the United States and its removal by precipitation, *Tellus*, *9*, 164-173, 1957

Kanakidou, M . and P J Crutzen, Scale problems in global tropospheric chemistry modeling Comparison of results obtained with a three-dimensional model, adopting longitudinally uniform and varying emissions of NO_x and NMHC, *Chemosphere*, *26*, 787-802, 1993

Kanakidou, M , and P J Crutzen, The photochemical source of carbon monoxide Importance, uncertainties and feedbacks, *Chemosphere. Global Change Sci* , *1*, 91-109, 1999

Kavouras, I , N Mihalopoulos, and E G Stephanou, Formation of atmospheric particles from organic acids produced by forests, *Nature*, *395*, 683-686, 1998

Kavouras, I , N Mihalopoulos, and E G Stephanou, Secondary organic aerosol formation vs primary organic aerosol emission In situ evidence for the chemical coupling between monoterpene acidic photooxidation products and new particle formation over forests, *Environ Sci Technol* , *33*, 1028-1037, 1999

Kerminen, V M , Roles of SO_2 and secondary organics in the growth of nanometer particles in the lower atmosphere, *J Aerosol Sci* , *30*, 1069-1078, 1999

Langner, J , and H Rodhe, A global three-dimensional model of the tropospheric sulfur cycle, *J Atmos Chem* , *13*, 225-263, 1991

Liousse, C , J E Penner, C Chuang, J J Walton, H Eddleman, and H Cachier, A global three-dimensional model study of carbonaceous aerosols, *J Geophys Res* , *101*, 19,411-19,432, 1996

Muller, J F , and G P Brasseur, IMAGES A three-dimensional chemical transport model of the global troposphere, *J Geophys Res* , *100*, 16,455-16,490, 1995

Newell, R E , J W Kidson, D G Vincent, and G.J Boer, *The General Circulation of the Tropical Atmosphere and Interactions with Extra Tropical Latitudes*, vol 2, MIT Press, Cambridge, Mass , 1974

Novakov, T , and C E Corrigan, Cloud condensation nucleus activity of the organic component of biomass smoke particles, *Geophys Res Lett* , *23*, 2141-2144, 1996

Novakov, T , and J E Penner, Large contribution of organic aerosols to cloud-condensation-nuclei concentrations, *Nature*, *365*, 823-826, 1993

Odum, J R , T Hoffmann, F Bowman, D Collins, R C Flagan, and J H Seinfeld, Gas/particle partitioning and secondary organic aerosol yields, *Environ Sci Technol* , *30*, 2580-2585, 1996

Olivier, J G J , et al , Description of EDGAR version 2 0 A set of global emission inventories of greenhouse gases and ozone depleting substances for all anthropogenic and most natural sources on a percountry basis on 1x1 grid, *RIVM Rep 771060 002 / TNO MEP Rep R96/119*, Neth Organ for Appl. Sci Res , The Hague, 1996

Pandis, S N , R. A Harley, G R Cass, and J H Seinfeld, Secondary organic aerosol formation and transport, *Atmos Environ. Part A*, *26*, 2269-2282, 1992

Poisson, N , M Kanakidou, and P J Crutzen, Impact of non-methane hydrocarbons on tropospheric chemistry and the oxidizing power of the global troposphere 3-dimensional modelling results, *J Atmos Chem in press*, 2000

Saxena, P , and L M Hildemann, Water-soluble organics in atmospheric particles A critical review of the literature and application of thermodynamics to identify candidate compounds, *J Atmos Chem* , *24*, 57-109, 1996.

Saxena, P , L Hildemann, P H Murry, and J H Seinfeld, Organics alter hydroscopic behavior of atmospheric particles, *J Geophys Res* , *100*, 18,755-18,770, 1995

Talbot, R W , M.O. Andreae, T W. Andreae, and R C Harriss, Regional aerosol chemistry of the Amazon Basin during the dry season, *J Geophys Res* , *93*, 1499-1508, 1988

Talbot, R W , M O Andreae, H Berresheim, P Artaxo, M Garstang, R. C Harriss, K M Beecher, and S M Li, Aerosol chemistry during the wet season in central Amazonia The influence of long-range transport, *J Geophys Res* , *95*, 16,955-16,969, 1990

Virkkula, A., R. Van Dingenen, F Raes, J Hjorth, Hydroscopic properties of aerosol formed by oxidation of limonene, *α*-pinene and *β*-pinene, *J Geophys. Res* , *104*, 3569-3579, 1999

Volz, A., and D. Kley, Evaluation of the Monsouris series of ozone measurements made at the nineteenth century, *Nature*, *332*, 240-242, 1988.

Wang, Y., D. J. Jacob, and J. A. Logan, Global simulation of tropospheric O_3-NO_x-hydrocarbon chemistry, 3, Origin of tropospheric ozone and effects of nonmethane hydrocarbons, *J. Geophys. Res.*, *103*, 10,757-10,768, 1998.

Yienger, J. J., and H. Levy II, Empirical model of global soil-biogenic NO_x emissions, *J. Geophys. Res.*, *100*, 11,447-11,464, 1995.

P. J. Crutzen, Atmospheric Chemistry Division, Max Planck Institute for Chemistry, P.O. Box 3060, 55022 Mainz, Germany (air@mpch-mainz.mpg.de)

F. J. Dentener, Institute for Marine and Atmospheric Research Utrecht, University of Utrecht, 3584 Utrecht, Netherlands. (F.J.Dentener@phys.uu.nl)

M. Kanakidou (corresponding author) and K. Tsigaridis, Environmental Chemical Processes Laboratory, Department of Chemistry, University of Crete, P.O. Box 1470, 71409 Heraklion, Greece (mariak@chemistry.uch.gr; tsigaridis@chemistry.uch.gr)

(Received August 19, 1999; revised November 23, 1999;
accepted November 30, 1999.)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9255–9264, APRIL 16, 2000

Secondary organics and atmospheric cloud condensation nuclei production

Veli-Matti Kerminen, Aki Virkkula, and Risto Hillamo
Finnish Meteorological Institute, Air Quality Research, Helsinki, Finland

Anthony S. Wexler
Department of Mechanical Engineering, University of Delaware, Newark

Markku Kulmala
Department of Physics, University of Helsinki, Helsinki, Finland

Abstract. The influence of secondary organics on atmospheric cloud condensation nuclei (CCN) production was investigated using a zero-dimensional box model that simulates the production of secondary organics in the gas phase, the transportation of these organics from gas to the particulate phase, and the resulting growth of the particles. Model simulations demonstrated that the growth of nanometer-size nuclei to a CCN size requires the presence of organics of extremely low volatility. These "nonvolatile" organics need to have saturation vapors pressures of the order of 0.01-0.1 parts per trillion or lower and, in order to induce sufficient nuclei growth, must have gas phase production rates of the order of 0.3-1 $\mu g\ m^{-3}\ d^{-1}$ under conditions typical for continental background areas. As the nuclei grow in size, they start to uptake volatile organics more efficiently. As a result, organic matter in both the nuclei grown into a CCN size and in the preexisting accumulation mode particles is expected to be dominated by "low-volatile" organics rather than organics that actually are responsible for the nuclei growth. The modeling results suggest that the monoterpene oxidation products identified so far in field or laboratory experiments, although important contributors to secondary particulate matter, are unlikely to be the ones that grow nuclei to a CCN size. In field experiments, positive identification of organics producing new CCN would require information on the chemical composition of particles smaller than about 0.1 μm in diameter, which is the size range where nonvolatile organics are likely to be enriched compared with other secondary or primary organics. Since the gas phase production rate of nonvolatile organics needs not to be very large in order to induce significant nuclei growth, more attention should also be paid to reaction products that have minor yields in smog chamber experiments.

1. Introduction

By influencing cloud properties and by providing a medium for many liquid phase reactions, particles able to act as cloud condensation nuclei (CCN) constitute an important subpopulation of atmospheric aerosol particles. A significant fraction of CCN are produced secondarily in the atmosphere, especially in remote locations where there are no major primary particle sources in the size range most critical to the CCN activation. Secondary CCN production involves two steps: (1) the formation of nanometer-size nuclei from gaseous precursors by homogeneous or ion-induced nucleation, and (2) the growth of these nuclei to sizes larger than about 50-80 nm in diameter. While neither of these two steps are well quantified at the moment, in current work we focus on examining the latter one. Whether particles grown to the potential CCN size range actually form cloud droplets depends on factors such as the meteorological conditions leading to cloud formation, the amount and properties of gaseous compounds condensing to droplets during the cloud formation, and the solubility, hygroscopicity, surface activity, and the mixing state of the material forming individual CCN. These issues have been investigated in a number of papers [*Kulmala et al.*, 1996; *Shulman et al.*, 1996; *Chuang et al.*, 1997; *Cruz and Pandis*, 1997, 1998; *Laaksonen et al.*, 1998; *Corrigan and Novakov*, 1999; *Facchini et al.*, 1999] and will not be considered here.

Candidate species contributing to secondary CCN production are sulfate and nitrate or their precursors, ammonia, and various organic compounds. The relative abundances of these secondary species in submicron aerosols

Copyright 2000 by the American Geophysical Union.

Paper number 1999JD901203.
0148-0227/00/1999JD901203$09.00

tell little about their actual role in CCN production, since the majority of CCN are too small to contribute significantly to submicron particulate mass. Some evidence exists that organics would be the primary class of compounds that determine the growth of nuclei to a CCN size. First, organic material has been observed to be enriched with respect to sulfate and other inorganics at particle sizes below 0.1 μm [*Novakov and Penner*, 1993; *Rivera-Carpio et al.*, 1996]. Second, both analyses and model studies have demonstrated that observed particle growth rates cannot usually be explained by the transportation of condensable inorganic vapors to these particles [*Marti et al.*, 1997; *Kerminen et al.*, 1997; *Kulmala et al.*, 1998a; *Weber et al.*, 1998]. The few direct observations of CCN production, limited thus far to forested areas, give also support to the important role of organic compounds during the nuclei growth [*Mäkelä et al.*, 1997; *Kavouras et al.*, 1998; *Leaitch et al.*, 1999].

Because of experimental difficulties in measuring the chemical composition of very small atmospheric particles, model studies investigating the influence of organic compounds on particle growth and CCN production are needed. Under atmospheric conditions, very few such studies have been performed. *Pandis et al.* [1993] investigated secondary organic formation in an urban atmosphere, and *Andrews et al.* [1997] simulated the time evolution particle size distribution in an air mass interacting with organic compounds emitted by the ocean and the forest. While highly informative in a qualitative sense, neither of these studies applied proper thermodynamic treatment of gas particle partitioning for secondary organic vapors.

In a preliminary investigation by *Kerminen* [1999], it was shown that only the very low volatile organic compounds can significantly contribute to the growth of nuclei in the atmosphere. In this work we investigate this topic in more detail by constructing a box model which explicitly treats the interaction of organic vapors with different size particles. The conditions for model simulations are chosen to be representative of those in continental background areas where atmospheric CCN production could be important because of the lack of major primary particle sources. The principal goal of this study is to provide insight into the following two questions: (1) which kind of properties are required for organics important to atmospheric CCN production, and (2) what requirements does this pose for the identification of these compounds in field or laboratory experiments? While unable to give quantitative link between CCN concentrations and the amount of secondary organic material in the atmosphere, our investigation is a step toward better understanding of the factors controlling atmospheric CCN production.

2. Approach

2.1. Gas-to-Particle Transfer of Organic Vapors

The transportation of organic compounds between the gas and particulate phase can be described using the condensation equation, according to which the flux of a compound i to a particle of diameter d_p is equal to [*Fuchs and Sutugin*, 1970]

$$I_i(d_p) = 2\pi D_i d_p \beta(Kn,\alpha_i)\left[C_i - Ke_i C_{eq,i}\right], \qquad (1)$$

$$\beta(Kn,\alpha_i) = \frac{1 + Kn}{1 + 0.377Kn + 1.33Kn(1 + Kn)/\alpha_i}. \qquad (2)$$

Here D_i is the gas phase diffusion coefficient of the compound i, C_i and $C_{eq,i}$ are its concentration in the gas phase and over a flat solution of the same composition as the particle, respectively. Ke is the Kelvin term, Kn is the particle Knudsen number, and α_i is the mass accommodation coefficient of the compound i on the particles surface. In case of large particles for which the Kelvin term is approximately equal to unity, the transportation of the compound i between the two phases continues until its gas phase concentration C_i becomes equal to $C_{eq,i}$. Consequently, $C_{eq,i}$ may also be called the equilibrium gas phase concentration for the compound i.

The value of $C_{eq,i}$, and thereby the distribution of the compound i between the gas phase and different-size particles, depends on whether the compound is adsorbed by the particle surface or absorbed by the organic matter in the particle [*Pankow*, 1994; *Venkatamaran et al.*, 1999]. For absorptive partitioning, $C_{eq,i}$ can be obtained from the following relation [*Bowman et al.*, 1997]:

$$C_{eq,i} = \frac{1}{K_{om,i}}\frac{m_i}{m_{om}} = \frac{1}{K_{om,i}}\frac{m_i}{f_{om}m_p}, \qquad (3)$$

where m_p is the particle mass, m_i is the total mass of the compound i in the particle, $m_{om}=f_{om}\times m_p$ is the total mass of the absorbing material (pure organics or a mixture of organics and inorganics such as water) in the particle, and $K_{om,i}$ ($m^3\ \mu g^{-1}$) is the absorption partitioning coefficient for the compound i. Theory predicts the following formula for $K_{om,i}$ [*Pankow*, 1994; *Odum et al.*, 1996]:

$$K_{om,i} = \frac{RT}{M_{om}\,p_{L,i}\,\gamma_{om,i}\times 10^6}, \qquad (4)$$

where R (8.206×10^{-5} m^3 atm mol^{-1} K^{-1}) is the ideal gas constant, T is the temperature, M_{om} (g mol^{-1}) is the mean molecular weight of the absorbing matter, $p_{L,i}$ (atm) is the saturation vapor pressure of the pure compound i as liquid (subcooled, if necessary), and $\gamma_{om,i}$ is its activity coefficient in the absorbing matter. Equation (4) demonstrates that the value of $K_{om,i}$ is governed mainly by the compound's saturation vapor pressure $p_{L,i}$. For a given compound $K_{om,i}$ varies with the temperature and the nature of the absorbing material in the particle (via $\gamma_{om,i}$).

The performance of the absorptive gas particle partitioning model described above has been demonstrated in a number of smog chamber experiments [*Odum et al.*, 1996, 1997; *Bowman et al.*, 1997; *Hoffmann et al.*, 1997; *Griffin et al.*, 1999], as well as in measurements carried out in the ambient air [*Liang and Pankow*, 1996; *Liang et al.*, 1997; *Simcik et al.*, 1998]. Adsorptive partitioning, the principal uptake mechanism of organic vapors for particles not having a pre-existing organic phase, is likely to be of less importance for the overall gas particle partitioning of organic matter in the atmosphere because it results typically in of most a few monolayers of the compound.

2.2. Model Description and Initial Conditions

The behavior of the system is simulated with a zero-dimensional box model representing the atmospheric boundary layer. Processes simulated inside the box include the production of secondary organics in the gas phase, transportation of these organics between the gas and the particulate phases, and the growth of particles due to the uptake of these compounds. Processes not taken into account are particle coagulation, deposition of particles and gaseous compounds onto the ground, and various heterogeneous reactions occurring on the surface or inside particles or cloud droplets. Neglecting coagulation and deposition is justified, since these two processes have little influence on the overall system behavior during the 24-hour simulation period applied. Heterogeneous reactions involving organic compounds might crucially influence the partitioning of organic matter between the gas phase and different size particles. At present, however, our understanding on heterogeneous organic chemistry is too limited to include it in the model.

Based on the above model assumptions, the concentration of a compound i in the gas phase is predicted using the following equation:

$$\frac{dC_i}{dt} = P_i - \int n(d_p)\, I_i(d_p)\, dd_p \,, \qquad (5)$$

where P_i is the gas phase production rate for the compound i, I_i is its flux to a single particle determined by (1)-(3), and $n(d_p)$ is the particle number size distribution. The particle number size distribution is assumed to be initially bimodal, with a lognormal nucleation mode centered at 5 nm and an accumulation mode centered at 150 nm (See Notation section). The selected accumulation mode number concentration corresponds to a submicron particulate mass of about 2 μg m^{-3}, a relatively low yet reasonable value for continental background areas.

In the model, both the nucleation and the accumulation mode is subdivided into 15 size sections which evolve in time independent of each other. The mean particle diameter of each section is calculated by keeping track of the amount of organic material condensing to particles in this section and by assuming a particle density of 1 g cm^{-3}. The condensation rate of organic vapors to particles in each section is calculated according to (1)-(3). Absorption of water by the particles is neglected. Considering the relatively low hygroscopic growth factors measured for organic particles [*Virkkula et al.*, 1999], this assumption is unlikely to cause an underestimate in particle sizes greater than a few percent at relative humidities below 90%. The accommodation coefficient of organic compounds on particles surfaces is assumed to be equal to unity. The molecular weight and the gas phase diffusion coefficient of the condensing organics are chosen to be equal to 150 g mol^{-1} and 0.1 cm^2 s^{-1}, respectively. An ambient temperature of 298 K is assumed.

For simplicity, both coarse and preexisting Aitken-mode particles are neglected in the base case simulation. Since these particles may act as a sink for secondary organic compounds, their potential influences will be returned to when performing the sensitivity analysis.

The organics identified in atmospheric aerosol particles, consisting mostly of alkanes, aldehydes, various acids, and certain multifunctional compounds, make up typically from 10 to 40% of the overall organic material in these particles [*Rogge et al.*, 1993; *Mazurek et al.*, 1997; *Seinfeld and Pandis*, 1998]. In the absence of detailed information on individual organic compounds and their properties in the particulate phase, we assume four generic groups of organics and term them as "nonvolatile" (NV), "low volatile" (LV), "semivolatile" (SV), and "highly volatile" (HV). Each group is treated as a single lumped species with an average $K_{om,i}$, as is done in many gas phase photochemical models.

The most important factors influencing the behavior of our system are the production rates of the organic compounds P_i and their gas particle partitioning coefficients $K_{om,i}$. The production rate of any secondary organic depends on the concentration of its precursors and oxidants (typically O_3 or the radicals OH or NO_3), and its yield from their reactions. Current gas phase chemistry models have quite limited capabilities in simulating the production of various condensable organics under different atmospheric conditions. In this work we assume that the production rates P_i are constant with time and equal to the values given in the Notation section. The selected production rates lead to the formation of about 3 μg m^{-3} of secondary particulate matter in our base case simulation, which can be considered a realistic amount in continental background areas [*Seinfeld and Pandis*, 1998].

Gas particle partitioning coefficients have been estimated for several tens of compounds identified in atmospheric aerosols. Obtained values for $K_{om,i}$ range from less than 10^{-5} m^3 ug^{-1} up to about 1 m^3 ug^{-1} [*Jang et al.*, 1997; *Liang et al.*, 1997; *Simcik et al.*, 1998]. A major source known to produce significant amounts of secondary particulate matter is the oxidation of biogenic hydrocarbons, especially monoterpenes. Smog chamber experiments involving these compounds have usually been described quite accurately using a two-component model, in which the less volatile component has $K_{om,i}$ somewhere between 0.01 and 1 m^3 ug^{-1} and the more volatile one between about 0.001 and 0.01 m^3 ug^{-1} [*Hoffmann et al.*, 1997; *Griffin et al.*, 1999]. Based on these observations, we select the values of 0.5 and 0.01 m^3 ug^{-1} for $K_{om,i}$ representing the low-volatile and semivolatile groups of organics, respectively. A value of 10^{-4} m^3 ug^{-1} is chosen for highly volatile organic compounds. This selection is somewhat arbitrary but, as will be shown later, does not affect significantly our system behavior. The value of $K_{om,i}$ for nonvolatile compounds is assumed to be infinitely large. Variation of $K_{om,i}$ with temperature or particle composition is not taken into account with base case but will be returned to when performing the sensitivity analysis.

3. Results and Discussion

3.1. General Behavior of the System

The evolution of the gas and particulate phases during our 24-hour base case simulation is illustrated in Figures 1-3. Figure 1 displays the modal mean diameter of the nucleation and the accumulation mode, along with the amount of secondary organic matter transferred into these modes. As we can see, the nucleation mode grows significantly in size, reaching a mean diameter of about 10 nm after 3 hours and a final diameter of 58 nm. The growth of the accumulation mode is less pronounced, the final mean diameter being equal to 274 nm. The nucleation mode particles become dominated rapidly by secondary organic matter, whereas only a few percent of the particulate matter in the accumulation mode is secondary during the first few simulation hours. Despite this, the vast majority of the secondary organics that are transferred to the particulate phase are in the accumulation mode.

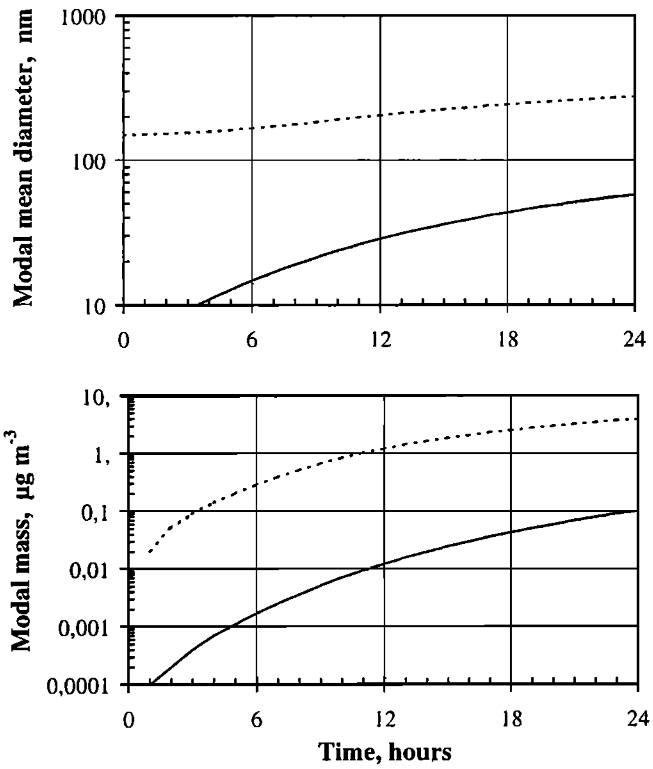


Figure 1. Time development of the nucleation (solid line) and the accumulation (dashed line) mode in the base case simulation. (top) The mode mean diameter, and (bottom) the amount of secondary organic matter transferred into these modes.

As mentioned earlier, secondary organics are assumed to consist of four different groups of compounds. The gas phase concentrations of the nonvolatile, low-volatile, and semivolatile groups of organics display quite different behavior in our system (Figure 2). The nonvolatile organics are transferred to the particulate phase over a timescale less than an hour, so their gas phase concentration do not exceed a few ng m^{-3}. Lower gas phase concentrations at later simulation times are due to increased total particle surface area. The gas phase concentration of both low-volatile and semivolatile organics increase with time. Compared with semivolatile organics, the lower volatility of the low-volatile group of organics makes its gas phase concentration significantly lower at later simulation times. The gas phase concentration of highly volatile organics not shown in Figure 2 follows closely that of the semivolatile group.

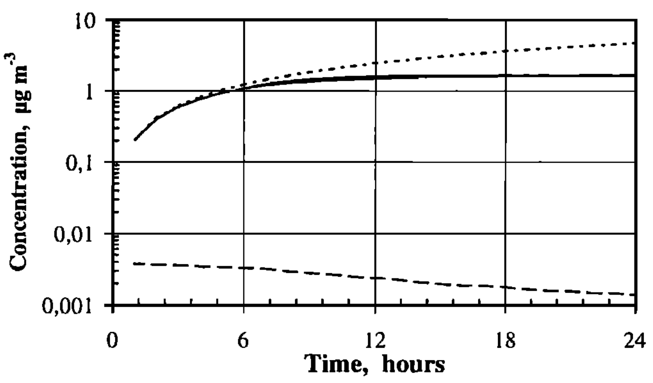


Figure 2. Gas phase concentrations of nonvolatile (dashed line), low-volatile (solid line), and semivolatile (dotted line) organic compounds during the base case simulation.

The fractional contribution of nonvolatile, low-volatile, and semivolatile organics to the overall organic matter in the two modes is shown in Figure 3. Initially, nonvolatile compounds are the dominant group of organics in both the modes. Reasons for this are the low overall concentration of organic matter in the particulate phase and the relatively low gas phase concentrations of more volatile organics. Both these factors disfavor the partitioning of volatile compounds to the particulate phase. As time evolves, the concentration of volatile compounds in the gas phase and their fractional contribution in the particulate phase increases. At the end of the simulation, the particulate phase is dominated by low-volatile organics followed by nonvolatile and semivolatile organics. The somewhat lower fraction of volatile organics in the nucleation mode compared with the accumulation mode is caused partially by mass transport limitations, and partially by the Kelvin term which enhances the vapor pressure of these compounds over the surface of nucleation mode particles. Highly volatile compounds are unimportant for the system behavior, since they constitute less than 0.1% of the organic matter in both the nucleation and the accumulation mode.

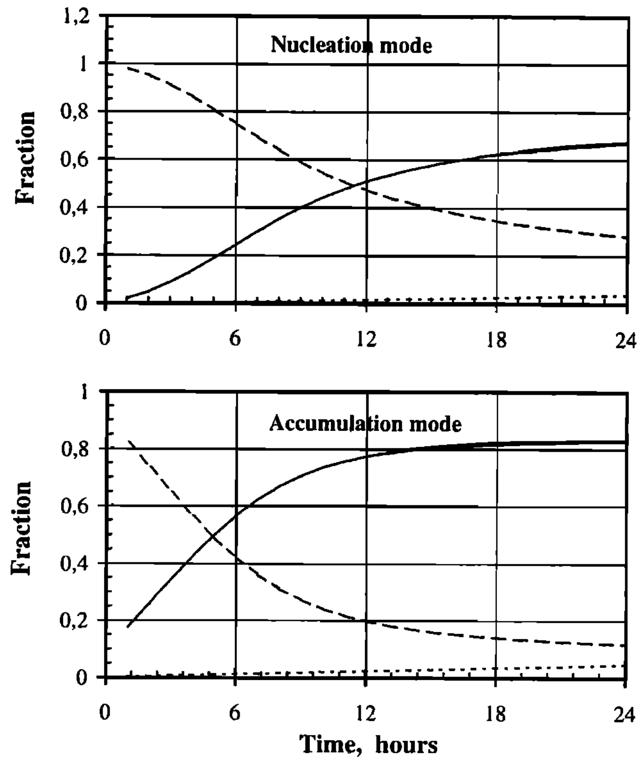


Figure 3. Fractional contribution of nonvolatile (dashed line), low-volatile (solid line), and semivolatile (dotted line) organic compounds to the overall organic matter transferred into the nucleation and the accumulation mode during the base case simulation.

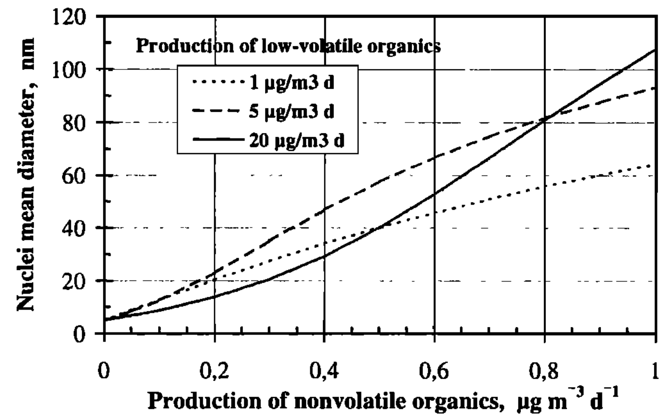


Figure 4. Mean diameter of the nucleation mode after 24 hours of simulation time as a function of the production rate for nonvolatile organics. The calculations are made for three different production rates (1, 5, and 20 µg m^{-3} d^{-1}) for low-volatile organics.

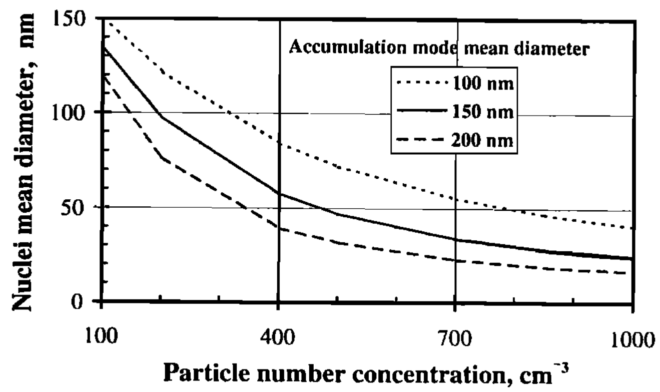


Figure 5. Mean diameter of the nucleation mode after 24 hours of simulation time as a function of accumulation mode particle number concentration and mean diameter. The range of values applied cover the submicron particulate matter concentration between about 0.1 and 10 µg m^{-3}.

In view of the above, it is clear that secondary organics with the least volatility are the most important ones when it comes to CCN production via the growth of nucleation mode particles. Figure 4 demonstrates how the final mean size of the nucleation mode depends on the production rate of nonvolatile and low-volatile organic compounds. From the figure we can immediately see that the incorporation of nonvolatile organics is a requisite for any significant growth of the nucleation mode. The reason is that without addition of new nonvolatile material to these particles, they rapidly become saturated with respect to more volatile compounds. At sufficiently high production rates for nonvolatile organics, the growth of the nucleation mode is assisted by the higher production rate for more volatile compounds. Contrary to this, at low production rates for nonvolatile organics the nuclei growth may even be suppressed by the presence of more volatile organics. In this latter case the volatile compounds enhance the accumulation mode surface area such that a significantly larger fraction of nonvolatile species will be transported to the accumulation mode instead of the nucleation mode.

The distribution of nonvolatile organics over the particulate phase is governed almost entirely by the available particle surface area. For more volatile organics there is less competition between different-size particles, since the partitioning of these compounds is driven by thermodynamic rather than kinetic factors. In the following subsection a more detailed analysis of the influence of various model parameters on our system behavior will be performed.

3.2. Sensitivity of the Nuclei Growth to Model Parameters

The primary uncertainties in our modeling results come from possible numerical errors caused by discretizations in time and particle size, as well as from uncertainties in the values of many model variables. The influences of numerical errors were examined by increasing the time and size resolution of the model, and they turned out to be negligible. The sensitivity of the system behavior to various model variables, within their expected range of variability under conditions typical for continental background areas, is investigated below.

3.2.1. Initial particle number size distribution. Preexisting particles affect the nuclei growth by providing an additional sink for secondary organic compounds. For nonvolatile organics, this sink term is proportional to the total particle surface area, whereas for more volatile organics it also depends on the chemical composition of the particulate phase. Figure 5 illustrates the influence of the accumulation mode mean diameter and particle number concentration on the nuclei growth. For larger particle surface areas (large N_A and $d_{p,A}$), a larger fraction of nonvolatile organics is transported to

Table 1. Mean Diameter of the Nucleation Mode After a 24-Hour Simulation Due to Changes in Different Model Parameters

Parameter	New value	$d_{p,N}$, nm
N_A, cm^{-3}; $d_{p,A}$, nm	100; 100	151
	100; 200	119
	200; 100	122
	200; 200	76
	1000; 100	41
N_N, cm^{-3}	1000; 200	17
	500	59
σ_A	5000	50
	1.3	62
α_{NV}	1.8	51
	0.3	45
	0.1	41
α_{NV}	0.01	26
	0.3	59
	0.1	62
σ_{surf}, N m^{-1} ; M_r, g mol^{-1}	0.01	75
	0.02; 100	88
$f_{om,init}$	0.07; 300	34
	0.05	56
	0.1	54
	0.2	50

the accumulation mode and the growth of nuclei is suppressed. If the accumulation mode surface area is sufficiently low (N_A< 150-300 cm^{-3} for organic vapor production rates assumed here), the nuclei may reach sizes of 100-150 nm after 24 hours of growth time. The simulations in Figure 5 were repeated by varying the geometric standard deviation of the accumulation mode. As expected, the nuclei grew somewhat faster when the accumulation mode was narrower. The difference in the nuclei mean diameter after the simulation was typically within 10-20% for geometric standard deviations between 1.3 and 1.8 (Table 1).

In addition to the accumulation mode, secondary organics may be transported to preexisting Aitken mode or coarse particles. The Aitken mode may dominate total particle surface area in air masses scavenged by rain and in air far from anthropogenic sources. Our simulations show that to influence the nuclei growth similar to the accumulation mode (N_A = 400 cm^{-3}, $d_{p,A}$ = 150 nm), the Aitken mode needs to have a particle number concentration of about 1000-1500 cm^{-3}, depending on the mean diameter and the geometric standard deviation of this mode. For coarse particles similar number-based estimates cannot be made, since the surface area of the coarse mode varies strongly with its width and mean diameter. Anyway, coarse particles could be a major sink for secondary organics in marine air masses having high sea-salt loadings, as well as in air masses containing a large amount of crustal matter.

3.2.2. Accommodation and diffusion coefficients. The rate at which any gaseous compound is transported to a particle is affected by its gas phase diffusion coefficient and its mass accommodation coefficient on the particle surface. For organic compounds, practically no experimental data on their mass accommodation coefficients α_i are available. We made a number of simulations in which the value of α_i for different organic groups was varied between 0.01 and 1 (Table 1). The influence of mass accommodation coefficient was most prominent for the nonvolatile group of organics, with their gas phase concentration increasing approximately as $1/\alpha_{NV}$ with the decrease in α_{NV}. The mean diameter of the nucleation mode decreased with decreasing α_{NV} . The gas phase concentration of the low-volatile group of organics increased slightly if the corresponding mass accommodation coefficient α_{LV} was decreased. This, together with the resulting lower accumulation mode surface area, induced a larger flux of organic compounds to the nucleation mode causing it to grow faster. Changing the mass accommodation coefficient of the more volatile organics had little influence on the system behavior.

The gas phase diffusion coefficient of a vapor is influenced greatly by its molecular weight. For light vapors such as water the diffusion coefficient in air is of the order 0.2 cm^2 s^{-1} under typical lower-tropospheric conditions, whereas for heavier vapors with molecular weights above 150 g mol^{-1} it is likely to be lower than 0.1 cm^2 s^{-1} assumed in our base case simulation [*Massman*, 1998]. We made additional simulations by varying the gas phase diffusion coefficients of organic compounds and by scaling their molecular masses accordingly based on the Grahams's law [*Massman*, 1998]. Changes in the system behavior, including the gas phase concentrations of the different organic groups and the mean diameters of the two modes, turned out to be negligible.

3.2.3. Kelvin effect. The partitioning of volatile compounds between the gas phase and particles smaller than about 0.1 μm in diameter is influenced by the Kelvin effect. In addition to the particle size, the magnitude of the Kelvin term is determined by the particle surface tension and the molecular weight of the compound condensing into the particle. The surface tension of pure organic particles can be as low as 0.02-0.03 N m^{-1}, whereas particles consisting mostly of water have surface tensions close to 0.07 N m^{-1} [*Shulman et al.*, 1996; *Seinfeld and Pandis*, 1998]. The molecular weights of secondary organics found in atmospheric aerosol particles are mostly in the range 150-250 g mol^{-1} [*Odum et al.*, 1996].

The influence of the Kelvin term on the nuclei growth is far from negligible (Table 1). Most of the variation in nuclei size is caused by the low-volatile group of organics which partition stronger to nuclei when the values of σ_{Surf} and M_i are lower. The amount of low-volatile organics in the nucleation mode varied by more than a factor 40 between the two extreme cases in Table 1, the amount of nonvolatile organics varied less than a factor 4.

3.2.4. Diurnal variability in gas phase production rates. The primary oxidants responsible for secondary organic vapor production have clear diurnal variability. The concentration of OH tracks closely the intensity of solar radiation, whereas that of NO_3 is almost negligible during the daytime. The concentration of O_3 peaks usually during daytime, even though its diurnal variability is substantially smaller than those of the OH or NO_3 radicals. To get some idea how variations in the oxidant concentrations might influence our system, we divided our simulation period into two parts representing the day (hours 06.00-18.00) and the night (hours 00.00-06.00 and 18.00-24.00). The production rates of organic compounds, although different between the day and the night, were assumed to be constant during the two periods.

To start with, it was assumed that the production of organic compounds is tied strongly to photochemical activity by selecting 10 times larger production rates during the day than during the night. The total amount of organics produced over the simulation period was kept the same as in our base case simulation. As one might expect, the gas phase concentrations of secondary organics differed from those in Figure 2 simulations. The concentration of nonvolatile organics remained low during the first simulation hours, increased to >7 ng m^{-3} rapidly after the "sunrise," decreased steadily during the day, and dropped rapidly <1 ng m^{-3} after the "sunset." The concentration of low-volatile organics increases monotonically but was delayed compared with the base case simulation because of the low production rates during the first simulation hours. Accumulation of organic compounds to the particulate phase occurred mainly during the daytime. Despite these differences, the mean diameters of both the nucleation and accumulation mode after the 24-hour simulation were almost the same as in our base case simulation.

While the overall production rate of secondary organics is probably higher during the daytime, it is likely that there are individual compounds produced mainly during the night as a result of the reaction of precursor species with NO_3. We performed a simulation in which the production rate of nonvolatile organics occurred mostly (90% of the overall production) during the night and that of more volatile organics mostly (90%) during the day. Note that although this choice may seem arbitrary, it is based on the observations that the reaction of certain monoterpenes with NO_3 might produce less volatile products than their reaction with O_3 or OH [*Hoffmann et al.*, 1997; *Griffin et al.*, 1999; *Hallquist et al.*, 1999]. If our simulation results are compared with those in the previous simulation, the principal differences are the slightly lower mean diameter of the nucleation mode (53 nm) after 24 hours of simulation time and the time evolution of nonvolatile organics in the gas phase with a peak concentration during the nighttime. In summary, although diurnal variations in the production rate of secondary organics may significantly influence their gas phase concentrations, the growth behavior of nuclei seems not to be affected substantially.

3.2.5. Other factors. In our base case simulation we assumed an initial organic fraction of 1% for both the nucleation and accumulation mode particles. In the ambient atmosphere this fraction could be substantially higher due to the presence of primary organic matter in the particulate phase or due to adsorption of gaseous organics by the particles surfaces. Simulations using different values for $f_{om,init}$ between about 0.01 and 0.2 did not change the system behavior very much (Table 1). We conclude that the growth of nuclei due to the uptake of secondary organics is not very sensitive to the initial organic content of these nuclei.

In our simulations we have assumed that the gas particle partitioning coefficient $K_{om,i}$ is constant with time for the each group of organic compounds. In reality, $K_{om,i}$ varies with time, since it is dependent on the chemical composition of the particulate phase, as well as on the ambient temperature and relative humidity [*Liang et al.*, 1997; *Jang and Kamens*, 1998, 1999]. Our simplified model cannot simulate the potential chances in $K_{om,i}$, caused by changes in the particle chemical composition such as its organic speciation and water content. The influence of diurnal temperature variations can be imitated in a rough way by changing the value of $K_{om,i}$ in a see-saw fashion with lower values during the day and higher values during the night when temperatures are lower. $K_{om,i}$ was assumed to reach maximum at 03.00 and minimum at 15.00. To alleviate comparisons to the base case simulation, the geometric mean of $K_{om,i}$ integrated over the simulation was kept unchanged.

As expected, diurnal variations in $K_{om,i}$ induced significant changes in the partitioning of volatile organics between the gas and particulate phases compared with the base case simulation. The changes were most evident when looking at the gas phase concentrations of different organics. When the diurnal variability in $K_{om,i}$ was within a factor of 20, changes in the mode mean diameter after the 24-hour simulation were <5% for the nucleation mode and <10% for the accumulation mode. For the nuclei, this kind of behavior is quite expected given the dominant role of nonvolatile organics with respect to their growth.

4. Organics Responsible for Atmospheric CCN Production

4.1. Required Properties

The simulations in the previous section demonstrate clearly that atmospheric CCN production caused by secondary organics is tied strongly to the presence of nonvolatile compounds in the system. In reality, even the least volatile compounds have a small saturation vapor pressure and thereby also a finite gas particle partitioning coefficient $K_{om,i}$. A question now emerges: how small can the value of $K_{om,NV}$ be without changing the nuclei growth behavior significantly from that when assuming $K_{om,NV}$ to be infinitely large? To answer this, we searched the minimum value of $K_{om,NV}$ for which the mean diameter of the nucleation mode after 24 hours of simulation time remained above 50 nm. This so-called critical gas particle partitioning coefficient $K_{om,crit}$, is plotted in Figure 6 as a function of the accumulation-mode particle number concentrations and for Kelvin terms of different magnitude. Other model parameters were kept equal to our base case simulation.

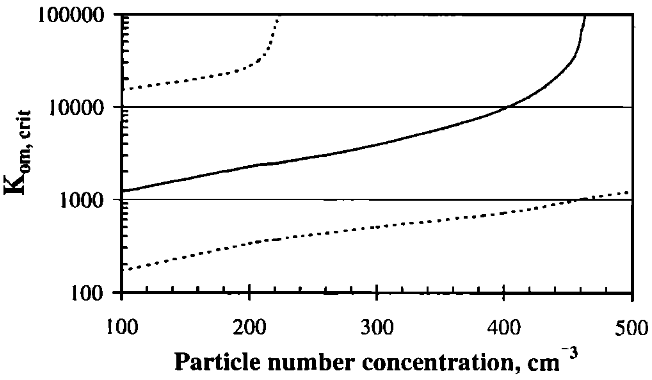


Figure 6. Critical gas particle partitioning coefficient as a function of accumulation mode particle number concentration. Solid line represents simulation with a Kelvin term similar to our base case simulation, whereas the bottom and top dashed lines are simulations in case of small (M_i= 100 g mol^{-1}, σ_{Surf}=0.03 N m^{-1}) and large (M_i= 300 g mol^{-1}, σ_{Surf}=0.07 N m^{-1}) Kelvin terms, respectively.

Figure 6 demonstrates that under conditions where atmospheric CCN production by secondary organics is viable, the gas particle partitioning coefficient of the nonvolatile group of organics needs to be of the order of 10^2-10^4 m^3 ug^{-1}. Because of the relatively small initial size of the nucleation mode particles, the exact value of $K_{om,crit}$ is extremely sensitive to the Kelvin term. This emphasizes the important role of the least volatile compounds, either organic or inorganic, in the very initial growth of nuclei formed in the atmosphere.

Knowing the range of values for $K_{om,crit}$, (4) offers a means to estimate the saturation vapor pressure $p_{L,i}$, required for an organic compound to make it important for atmospheric nuclei growth. To apply (4), one needs to know the molar mass of the absorbing phase M_{om} and the activity coefficient of the organic compound in it $\gamma_{om,i}$. The values of M_{om} for atmospheric particles are expected to be in the range 100-300 g mol^{-1} [*Pankow*, 1994; *Odum et al.*, 1996]. Measured activity coefficients vary typically between about 1 and 5, even though values in excess of 10 have also been reported for certain compounds [*Jang et al.*, 1997; *Jang and Kamens*, 1998]. With this information an upper-limit estimate of about 1 part per trillion (ppt) can be obtained for $p_{L,i}$ ($K_{om,i}$ =100, $\gamma_{om,i}$=1). By taking into account that $\gamma_{om,i}$ is somewhat larger than unity and that the required values of $K_{om,i}$ probably are larger than 100, we may estimate that in practice the saturation vapor pressure of an organic compound needs to be about 0.01-0.1 ppt or lower to make it important for the growth of nuclei in the atmosphere. A vapor pressure of similar magnitude was found by *Kulmala et al.* [1998a] using a different approach, in which observed particle growth times in a boreal forest were analyzed with model simulations.

The vast majority of organics identified in atmospheric aerosol particles have a saturation vapor pressure of the order of several parts per billion (ppb) or greater [*Jang et al.*, 1997; *Liang et al.*, 1997; *Simcik et al.*, 1998]. Based on our earlier classification, these compounds belong to the semivolatile or highly volatile group of organics. The oxidation of terpenoids, especially monoterpenes, have been shown to produce organics of relatively low volatility [*Galogirou et al.*, 1999; *Griffin et al.*, 1999; *Yu et al.*, 1999a]. Of those identified tentatively in smog chamber experiments, the least volatile ones include hydroxy pinonic acid, pinic acid, hydroxy pinonaldehydes, pinonic acid, and nor-pinonic acid and its isomers [*Hoffmann et al.*, 1998; *Jang and Kamens*, 1999; *Yu et al.*, 1999a]. The saturation vapor pressures of these compounds have been estimated to range from a few ppt to several tens of ppb, depending on the compound and the atmospheric conditions [*Jang and Kamens*, 1999; *Yu et al.*, 1999a]. According to our classification, these compounds belong to the low-volatile or semivolatile groups of organics and, although important contributors to secondary particulate matter in the atmosphere, they are unlikely to be the ones that actually grow nuclei to a CCN size.

Laboratory experiments concerning the oxidation of monoterpenes and isoprene have been observed to produce minor fractions of several yet unidentified multifunctional compounds [*Wängberg et al.*, 1997; *Chien et al.*, 1998; *Hoffmann et al.*, 1998; *Hallquist et al.*, 1999; *Jang and Kamens*, 1999; *Yu et al.*, 1999a]. If having sufficiently low saturation vapor pressures, it is quite possible that some of these compounds could belong to the nonvolatile group of organics necessary for atmospheric CCN production. In the next section we briefly discuss the implications of our modeling results for identifying these nonvolatile organics in field or laboratory experiments.

4.2. Implications for Experimental Work

In previous sections we outlined what kind of properties are required for secondary organics that could be important for atmospheric CCN production. While the modeling results were dependent on several parameters that may vary greatly with varying atmospheric conditions, there are many general conclusions that probably hold and thus can be made use of in experiments carried out in the field and laboratory.

We demonstrated that the growth of nanometer size nuclei to a CCN size requires the presence of rather nonvolatile organics with saturation vapor pressure of the order 0.01-0.1 ppt or lower. This is quite a severe constraint, limiting the candidate species to multifunctional compounds that condense rapidly to the particulate phase. The low vapor pressure of these compounds certainly poses limitations with the regard of their detection using many current analyzing techniques.

In systems where atmospheric CCN production via secondary organics is viable, the production rate of nonvolatile organics is expected to lie in the range 0.1-1 μg m^{-3} d^{-1}. This rate may be substantially lower than the overall production rate of secondary organics in the system. When noting further that the nonvolatile group of organics may consists of a number of different compounds, it becomes clear that even compounds that have relatively low yields in precursor oxidation reactions could contribute significantly to atmospheric CCN production. This is worth keeping in mind when interpreting smog chamber experiments.

In recent field experiments carried out in forested areas, a number of different terpene oxidation products have been identified in the particulate phase [*Yu et al.*, 1999b; *Kavouras et al.*, 1998, 1999]. Particle collection in these experiments has been based either on bulk filter sampling or on impactor sampling with only one or two size fractions in the submicron size range. Our model results show that the nonvolatile organics responsible for atmospheric CCN production are expected to be enriched in very small particles but, despite this, may constitute less than 10% of the overall submicron organic mass. This means that any collection device unable to collect separately particles smaller than some 0.1 μm in diameter cannot give reliable estimates on whether the measured compounds really are the ones that produce new CCN or whether they just partition into preexisting accumulation mode particles.

The model runs demonstrated further that the overall gas phase concentration of nonvolatile organics is unlikely to exceed a few ng m^{-3} in the atmosphere. Thus any compound observed at greater abundances in the gas phase is unlikely to belong to the group of compounds that are responsible for atmospheric CCN production.

5. Summary and Conclusions

Atmospheric CCN production caused by secondary organic compounds was investigated using a zero-dimensional box model that simulates aerosol dynamics coupled with simplified gas phase chemistry. Based on their gas particle partitioning properties, the secondary organic compounds were divided into four groups termed nonvolatile, low-volatile, semivolatile, and highly volatile organics. Gas phase production rates of these organics were prescribed. Transportation of organic compounds between the gas phase and different size particles were modeled explicitly using a condensation equation and a proper thermodynamic treatment for gas particle partitioning. The initial particle number size distribution was assumed to be bimodal with an accumulation mode centered at 150 nm and a nucleation mode at 5 nm.

Model simulations demonstrated that the growth of nanometer size nuclei to a CCN size requires the presence of organics of extremely low volatily. These nonvolatile organics need to have saturation vapors pressures of the order of 0.01-0.1 ppt or lower and, in order to induce sufficient nuclei growth, must have gas phase production rates of the order of 0.3-1 μg m^{-3} d^{-1} under conditions typical for continental background areas. As the nuclei grow in size, they start to uptake volatile organics more efficiently. As a result, organic matter in both the nuclei grown into a CCN size and in the preexisting accumulation mode particles is expected to be dominated by low-volatile organics rather than organics that actually are responsible for the nuclei growth.

In the atmosphere, direct observations of secondary CCN production caused by secondary organics have been observed mainly in forested areas [*Mäkelä et al.*, 1997; *Kavouras et al.*, 1998; *Kulmala et al.*, 1998b; *Leaitch et al.*, 1999]. Forests are known emitters of monoterpenes and other terpenoids capable of forming secondary particulate matter. According to our modeling results, the monoterpene oxidation products identified so far, although important contributors to secondary particulate matter, seem too volatile to be the ones that grow small nuclei to a CCN size. In field experiments, positive identification of organics responsible for atmospheric CCN production would require information on the chemical composition of particles smaller than about 0.1 μm in diameter, which is the size range where nonvolatile organics are likely to be enriched compared with other secondary or with primary organics. Since the gas phase production rate of nonvolatile organics needs not to be very large in order to induce significant nuclei growth, more attention should also be paid to reaction products that have minor yields in smog chamber experiments.

Our model exercise assumed that secondary organics found in the particulate phase were produced initially by various gas phase reactions in the atmosphere. In addition to this formation pathway, secondary organics could be produced by heterogeneous reactions occurring on the surface or inside aerosol particles or in cloud droplets. It is clear that reactions in cloud droplets do not assist secondary CCN production as they involve particles that already are active as CCN. Instead, inhibition of the nuclei growth due to the consumption of secondary organics or their precursors by the cloud water is quite possible and should be taken into account in later studies related to this subject. Reactions occurring on the nuclei surfaces could produce nonvolatile organics that are not evaporated back to the gas phase. In case these reactions would prefer nuclei from preexisting accumulation mode particles, even lower productions rates of nonvolatile organics than predicted by our model simulations would be sufficient to induce secondary CCN formation. Unfortunately, the lack of proper mechanistic undertanding on heterogeneous organic chemistry precludes us to make more definite conclusions on whether heterogenous reactions could play some role in atmospheric CCN production.

Finally, it is worth noting that organics having a very low volatility may not only condense on preexisting particles but also produce new nanometer size particles. If consuming a significant fraction of low-volatility organics from the gas phase, the production of new particles could compete with the growth of larger nuclei. A detailed analysis of this subject goes beyond the scope of this paper, even though the very low mass concentrations of newly formed particles compared with larger particles lead us to believe that this kind of competition probably is quite rare under atmospheric conditions.

Notation

N_N	particle number concentration, 1000 cm^{-3}.
$d_{p,N}$	mode mean diameter, 5 nm.
σ_N	mode geometric standard deviation, 1.3.
N_A	particle number concentration, 400 cm^{-3}
$d_{p,A}$	mode mean diameter, 150 nm.
σ_A	mode geometric standard deviation, 1.5.
T	ambient temperature, 298 K.
$f_{om,init}$	initial particle organic fraction, 0.01.
ρ_p	particle density, 1 g cm^{-3}.
σ_{Surf}	particle surface tension, 0.05 N m^{-1}.
α_i	species accommodation coefficient, 1.
D_i	species gas phase diffusion coefficient, 0.1 cm^2 s^{-1}.
M_i	species molecular weight, 150 g mol^{-1}.
P_{NV}	production rate of nonvolatile organics, 0.5 μg m^{-3} d^{-1}.
P_{LV}	production rate of low-volatile organics, 5 μg m^{-3} d^{-1}.
P_{SV}	production rate of semivolatile organics, 5 μg m^{-3} d^{-1}.
$K_{om,NV}$	absorbtion partitioning coefficient (nonvolatile), ∞.
$K_{om,LV}$	absorbtion partitioning coefficient (low volatile), 0.5 m^3 μg^{-1}.
$K_{om,SV}$	absorbtion partitioning coefficient (semivolatile), 0.01 m^3 μg^{-1}.
$K_{om,H}$	absorbtion partitioning coefficient (highly volatile), 10^{-4} m^3 μg^{-1}.

Acknowledgments. This work has been funded by the Academy of Finland and by the Maj and Tor Nessling foundation.

References

Andrews, E., S. M. Kreidenweis, J. E. Penner, and S. M. Larson, Potential origin of organic cloud condensation nuclei observed at marine site, *J. Geophys. Res.*, *102*, 21,997-22,012, 1997.

Bowman, F. M., J. R. Odum, J. H. Seinfeld, and S. N. Pandis, Mathematical model for gas-particle partitioning of secondary organic aerosols, *Atmos. Environ.*, *31*, 3921-3931, 1997.

Chien, C.-J., M. J. Charles, K. G. Sexton, and H. E. Jeffries, Analysis of airborne carboxylic acids and phenols as their pentafluorobenzyl derivatives: Gas chromatography/ion trap mass spectrometry with a novel chemical ionization reagent, PFBOH, *Environ. Sci. Technol.*, *32*, 299-309, 1998.

Chuang, P. Y., R. J. Charlson, and J. H. Seinfeld, Kinetic limitations on droplet formation in clouds, *Nature*, *390*, 594-596, 1997.

Corrigan, C. E., and T. Novakov, Cloud condensation nucleus activity of organic compounds: A laboratory study, *Atmos. Environ.*, *33*, 2661-2668, 1999.

Cruz, C., and S. N. Pandis, A study of the ability of pure secondary organic aerosol to act as cloud condensation nuclei, *Atmos. Environ.*, *31*, 2205-2214, 1997.

Cruz, C., and S. N. Pandis, The effect of organic coatings on the cloud condensation nuclei activation of inorganic atmospheric aerosol, *J. Geophys. Res.*, *103*, 13,111-13,123, 1998.

Facchini, M. C., M. Mircea, S. Fuzzi, and R. J. Charlson, Cloud albedo enhancement by surface-active organic solutes in growing droplets, *Nature*, *401*, 257-259, 1999.

Fuchs, N. A., and A. G. Sutugin, *Highly Dispersed Aerosols*, Butterworth-Heinemann, Newton, Mass., 1970.

Galogirou, A., B. R. Larsen, and D. Kotzias, Gas-phase terpene oxidation products: A review, *Atmos. Environ.*, *33*, 1423-1439, 1999.

Griffin, R. J., D. R. Cocker III, R. C. Flagan, and J. H. Seinfeld, Organic aerosol formation from the oxidation of biogenic hydrocarbons, *J. Geophys. Res.*, *104*, 3555-3567, 1999.

Hallquist, M., I. Wängberg, E. Ljungström, I. Barnes, and K.-H. Becker, Aerosol and product yields from NO_3 radical-initiated oxidation of selected monoterpenes, *Environ. Sci. Technol.*, *33*, 553-559, 1999.

Hoffmann, T., J. R. Odum, F. Bowman, D. Collins, D. Klockow, R. C. Flagan, and J. Seinfeld, Formation of organic aerosols from the oxidation of biogenic hydrocarbons, *J. Atmos. Chem.*, *26*, 189-222, 1997.

Hoffmann, T., R. Bandur, U. Marggraf, and M. Linscheid, Molecular composition of organic aerosols formed in the α-pinene/O_3 reaction: Implications for new particle formation processes, *J. Geophys. Res.*, *103*, 25,569-25,578, 1998.

Jang, M., and R. M. Kamens, A thermodynamic approach for modeling partitioning of semivolatile organic compounds on atmospheric particulate matter: Humidity effects, *Environ. Sci. Technol.*, *32*, 1237-1243, 1998.

Jang, M., and R. M. Kamens, Newly characterized products and composition of secondary aerosols from the reaction of α-pinene with ozone, *Atmos. Environ.*, *33*, 459-474, 1999.

Jang, M., R. M. Kamens, K. B. Leach, and M. R. Strommen, A thermodynamic approach using group contribution methods to model the partitioning of semivolatile organic compounds on atmospheric particulate matter, *Environ. Sci. Technol.*, *31*, 2805-2811, 1997.

Kavouras, I. G., N. Mihalopoulos, and E. G. Stephanou, Formation of atmospheric particles from organic acids produced by forests, *Nature*, *395*, 683-686, 1998.

Kavouras, I. G., N. Mihalopoulos, and E. G. Stephanou, Formation and gas/particle partitioning of monoterpenes photooxidation products over forests, *Geophys. Res. Lett.*, *26*, 55-58, 1999.

Kerminen, V.-M., Roles of SO_2 and secondary organics in the growth of nanometer particles in the lower atmosphere, *J. Aerosol Sci.*, *30*, 1069-1078, 1999.

Kerminen, V.-M., A. S. Wexler, and S. Potukuchi, Growth of freshly nucleated particles in the troposphere: Roles of NH_3, H_2SO_4, HNO_3, and HCl, *J. Geophys. Res.*, *102*, 3715-3724, 1997.

Kulmala, M., P. Korhonen, T. Vesala, H.-C. Hansson, K. Noone, and B. Svenningsson, The effect of hygroscopicity on cloud droplet formation, *Tellus, Ser. B*, *48*, 347-360, 1996.

Kulmala, M., A. Toivonen, J. M. Mäkelä, and A. Laaksonen, Analysis of the growth of nucleation mode particles observed in boreal forest, *Tellus, Ser. B*, *50*, 449-462, 1998a.

Kulmala, M., J. M. Mäkelä, I. Koponen, and L. Pirjola, Formation of cloud condensation nuclei in boreal forest area, *J. Aerosol Sci.*, *29*, suppl. 1, S567-S568, 1998b.

Laaksonen, A., P. Korhonen, M. Kulmala, and R. J. Charlson, Modification of the Köhler equation to include soluble trace gases and slightly soluble substances, *J. Atmos. Sci.*, *55*, 853-862, 1998.

Leaitch, W. R., J. W. Bottenheim, T. A. Biesenthal, S.-M. Li, P. S. K. Liu, K. Asalian, H. Dryfhout-Clark, and F. Hopper, A case study of gas-to-particle conversion in an eastern Canadian forest, *J. Geophys. Res.*, *104*, 8095-8111, 1999.

Liang, C., and J. F. Pankow, Gas/particle partitioning of organic compounds to environmental tobacco smoke: Partition coefficient measurements by desorption and comparison to urban particulate matter, *Environ. Sci. Technol.*, *30*, 2800-2805, 1996.

Liang, C., J. F. Pankow, J. R. Odum, and J. H. Seinfeld, Gas/particle partitioning of semivolatile organic compounds to model inorganic, organic, and ambient smog aerosols, *Environ. Sci. Technol.*, *31*, 3086-3092, 1997.

Mäkelä, J. M., P. Aalto, V. Jokinen, T. Pohja, A. Nissinen, S. Palmroth, T. Markkanen, K. Seitsonen, H. Lihavainen, and M. Kulmala, Observations of ultrafine aerosol particle formation and growth in boreal forest, *Geophys. Res. Lett.*, *24*, 1219-1222, 1997.

Marti, J. J., R. J. Weber, P. H. McMurry, F. Eisele, D. Tanner, and A. Jefferson, New particle formation at a remote continental site: Assessing the contributions of SO_2 and organic precursors, *J. Geophys. Res.*, *102*, 6331-6339, 1997.

Massman, W. J., A review of the molecular diffusivities of H_2O, CO_2, CH_4, CO, O_3, SO_2, NH_3, N_2O, NO, and NO_2 in air, O_2 and N_2 near STP, *Atmos. Environ.*, *32*, 1111-1127, 1998.

Mazurek, M., M. C. Masonjones, H. D. Masonjones, L. G. Salmon, G. R. Cass, K. A. Hallock, and M. Leach, Visibility-reducing organic aerosols in the vicinity of Grand Canyon National Park: Properties observed by high-resolution gas chromatography, *J. Geophys. Res.*, *102*, 3779-3793, 1997.

Novakov, T., and J. E. Penner, Large contribution of organic aerosols to cloud-condensation nuclei concentrations, *Nature*, *365*, 823-826, 1993.

Odum, J. R., T. Hoffmann, F. Bowman, D. Collins, R. Flagan, and J. H. Seinfeld, Gas/particle partitioning and secondary organic aerosol yields, *Environ. Sci. Technol.*, *30*, 2580-2585, 1996.

Odum, J. R., T. P. W. Jungkamp, R. J. Griffin, H. J. L. Forstner, R. C. Flagan, and J. H. Seinfeld, Aromatics, reformulated gasoline, and atmospheric organic aerosol formation, *Environ. Sci. Technol.*, *31*, 1890-1897, 1997.

Pandis, S. N., A. S. Wexler, and J. H. Seinfeld, Secondary organic aerosol formation and transport, II, Predicting the ambient secondary organic aerosol size distribution, *Atmos. Environ.*, *27A*, 2403-2416, 1993.

Pankow, J. F., An absorption model of gas/particle partitioning of organic compounds in the atmosphere, *Atmos. Environ.*, *28*, 185-188, 1994.

Rivera-Carpio, C. A., C. E. Corrigan, T. Novakov, J. E. Penner, C. R. Rogers, and J. C. Chow, Derivation of contributions sulfate and carbonaceous aerosols to cloud condensation nuclei from mass size distributions, *J. Geophys. Res.*, *101*, 19,483-19,493, 1996.

Rogge, W. F., M. A. Mazurek, L. M. Hildemann, G. R. Cass, and B. R. T. Simoneit, Quantification of urban organic aerosols at a molecular level: Identification, abundance and seasonal variation, *Atmos. Environ.*, *27A*, 1309-1330, 1993.

Seinfeld, J. H., and S. N. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, John Wiley, New York, 1998.

Shulman, M. L., M. C. Jacobson, R. J. Charlson, R. E. Synovec, and T. E. Young, Dissolution behavior and surface tension effects of organic compounds in nucleating cloud droplets, *Geophys. Res. Lett.*, *23*, 277-280, 1996.

Simcik, M. F., T. P. Franz, H. Zhang, and S. J. Eisenreich, Gas-particle partitioning of PCBs and PAHs in the Chicago urban and adjacent coastal atmophere: States of equilibrium, *Environ. Sci. Technol.*, *32*, 251-257, 1998.

Venkatamaran, C., S. Thomas, and P. Kulkarni, Size distributions of polycyclic aromatic hydrocarbons: Gas/particle partitioning to urban aerosols, *J. Aerosol Sci.*, *30*, 759-770, 1999.

Virkkula, A., R. Van Dingenen, F. Raes, and J. Hjorth, Hygroscopic properties of aerosol formed by oxidation of limonene, α-pinene, and β-pinene, *J. Geophys. Res.*, *104*, 3569-3579, 1999.

Wängberg, I, I. Barnes, and K. H. Becker, Product and mechanistic study of the reaction of NO_3 radicals with α-pinene, *Environ. Sci. Technol.*, *31*, 2130-2135, 1997.

Weber, R. J., P. H. McMurry, L. Mauldin, D. J. Tanner, F. L. Eisele, F. J. Brechtel, S. M. Kreidenweis, G. L. Kok, R. D. Schillawski, and D. Baumgardner, A study of new particle formation and growth involving biogenic and trace gas species measured during ACE 1, *J. Geophys. Res*, *103*, 16,385-16,396, 1998.

Yu, J., D. R. Cocker III, R. J. Griffin, R. C. Flagan, and J. H. Seinfeld, Gas-phase ozone oxidation of monoterpenes: Gaseous and particulate products, *J. Atmos. Chem.*, *34*, 207-258, 1999a.

Yu, J., R. J. Griffin, D. R. Cocker III, R. C. Flagan, and J. H. Seinfeld, Observation of gaseous and particulate products of monoterpene oxidation in forest atmospheres, *Geophys. Res. Lett.*, *26*, 1145-1148, 1999b.

R. E. Hillamo, V.-M. Kerminen (corresponding author), A. Virkkula, Finnish Meteorological Institute, Air Quality Research, Sahaajankatu 20E, FIN-00810 Helsinki, Finland. (veli-matti.kerminen@fmi.fi)

M. Kulmala, Department of Physics, University of Helsinki, P.O. Box 9, FIN-00014. Helsinki, Finland.

A. S. Wexler, Department of Mechanical Engineering, University of Delaware, Newark, DE 19716.

(Received August 26, 1999; revised November 17, 1999; accepted December 21, 1999)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9265–9282, APRIL 16, 2000

Exploratory analysis of the distribution of condensation nuclei in the Northern Hemisphere upper troposphere and lower stratosphere during the late 1970s

Andrew G. Detwiler and L. Ronald Johnson

Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, Rapid City

Allison G. Schauer

Science and Engineering Services, Inc., NASA Wallops Flight Facility, Wallops Island, Virginia

Abstract. Data obtained from the Global Atmospheric Sampling Program (GASP) during the late 1970s are used to evaluate the hypothesis that particulate emissions by aircraft are a significant source for the upper tropospheric/lower stratospheric fine and ultrafine aerosol. Measurements of condensation nuclei (CN) were acquired above 6 km mean sea level as part of GASP from October 1977 through June 1979, using automated packages on up to four commercial airliners simultaneously flying their normal daily routes. These routes spanned much of the midlatitude Northern Hemisphere, with the notable exceptions of the former Soviet Union and China. The GASP measurements are analyzed to determine the vertical, geographical, and seasonal distribution of upper tropospheric/lower stratospheric CN in clear air, spanning an altitude range from several kilometers above to several kilometers below the tropopause. The general lack of vertical gradients in median CN mixing ratio at altitudes in the troposphere suggests that CN in the clear troposphere are generally well mixed. CN mixing ratios decrease with distance above the tropopause, consistent with a tropospheric source for stratospheric CN. Median CN mixing ratios for selected regions were compared to the Northern Hemisphere median, by altitude and season. Regional median CN mixing ratios can differ from the hemispheric median by more than a factor of 2 but never an order of magnitude. The lack of vertical variation of CN concentrations in the troposphere is not at all consistent with the vertical variation of fuel consumed by aircraft, which peaks at altitudes near the tropopause. The observations of CN show some regional medians exceeding Northern Hemisphere medians in some seasons and not others, but the regions with the highest medians are not always those with the highest aircraft fuel consumption. We conclude that aircraft emissions are not a major component of the global clear air background CN measured during GASP.

1. Introduction

Anthropogenic aircraft emissions are of special interest, despite their relatively small contribution to the global anthropogenic emission of pollutants. Most aircraft emissions occur high in the troposphere where removal mechanisms are less effective than nearer the ground where most other anthropogenic emissions occur. The aircraft emissions on which this report will focus are emissions of ultrafine and fine particulate matter ($d \leq 0.5$ μm) produced during combustion in aircraft engines. Reports of recent measurements of this particulate matter during the Subsonic Aircraft: Contrail and Cloud Effects Special Study (SUCCESS) are surveyed by *Toon and Miake-Lye* [1998]. These include *Miake-Lye et al.* [1998], *Hagen et al.* [1998], *Pueschel et al.* [1998], *Anderson et al.* [1998a,b], *Dibb et al.* [1998], and *Paladino et al.* [1998]. Earlier measurements are reported by, for example, *Brasseur et al.* [1998], *Schlager et al.* [1997], *Schumann et al.* [1996], *Fahey et al.* [1995], *Spicer et al.* [1994], *Whitefield et al.* [1993], *Spicer et al.* [1992], and *Rosen and Hoffman* [1977]. These reports demonstrate that aircraft engine exhaust particles are composed mainly of carbonaceous soot, sulfates, oxides of the metals of which the engines are constructed, oxides of trace components of the fuels, and components of ambient aerosol particles processed through the aircraft engines. The concentration of particles volatile at temperatures above ~150°C is about twice that of particles that do not volatilize at such temperatures. The volatile particles tend to be from the small end of the particle size spectrum, which ranges from a few nanometers to several micrometers.

The fate of aircraft particulate emissions differs from that of other anthropogenic particulate emissions due to the height of release. Most anthropogenic pollutant sources are at or near the surface, and material emitted there may or may not be transported upward out of the surface boundary layer. In contrast, aircraft exhaust is released into the atmosphere over a range of altitudes, with most emissions during cruise flight in the upper troposphere. At cruise altitudes (typically 9-13 km mean sea level (msl)) where most emissions occur, the exhaust is injected directly into the upper troposphere and lower stratosphere. In these regions, removal mechanisms, such as scavenging by cloud particles, operate in relatively smaller volumes of the atmosphere, and for shorter periods, compared to these same mechanisms operating in the lower troposphere. Removal of exhaust particulate from the upper troposphere and lower stratosphere thus requires a longer timescale than at lower altitudes.

In the summary compiled by *Friedl* [1997], virtually all of the black carbon aerosol in the upper troposphere/lower stratosphere is attributed to aircraft emissions, although black carbon aerosol typically represents a negligible (<<1%) fraction of the total aero-

Copyright 2000 by the American Geophysical Union.

Paper number 2000JD900022.
0148-0227/00/2000JD900022$09.00

sol mass in these regions. Increases in concentration of cloud condensation nuclei (CCN, a subset of the condensation nuclei population that serve as nuclei for condensation of water at supersaturations reached in natural clouds), due to aircraft emissions, of from 5 to 500 cm^{-3} are estimated for heavily traveled flight corridors. The global impact of aircraft soot emissions on the atmospheric energy balance is estimated to be small relative to the radiative impact of aircraft NO_x emissions (NO_x emissions lead to increases in O_3, a greenhouse gas) and CO_2 emissions (CO_2 is another greenhouse gas).

Direct observations of the impact of aircraft on concentrations of atmospheric aerosol particulates are limited. *Hofmann* [1991] found a correlation between the increase in lower stratospheric sulfate mass loading (~5% / yr) and the increase in jet fuel consumption in the stratosphere (~5% / yr) from the 1970s into the 1990s and postulated that emission of sulfur oxides by aircraft led to the increase in atmospheric aerosol sulfate. *Hofmann* [1993] noted that this increase was mainly at the 15-20 km altitude level over Laramie, Wyoming (corresponding to lower stratospheric regions), with decreases at lower levels and weaker trends above. *Hoffman et al.* [1998] concluded that aircraft particulate emissions over the Laramie, Wyoming, region increased the background upper tropospheric aerosol concentration by about 10%. *Strom and Ohlsson* [1998], using in situ aerosol and cloud particle measurements, identified regions within cirrus cloud formations where aircraft particulate emissions have apparently resulted in a higher concentration of cloud ice particles.

Baughcum et al. [1996a,b] constructed an extensive emission inventory for civil aircraft. Fossil fuel consumption by aircraft constitutes ~ 2% of the total fossil fuel combusted globally each year in recent decades [*Baughcum et al*, 1996b; *Brown et al.*, 1997, *Friedl*, 1997]. A small seasonal variation in global aircraft fuel consumption was found with a peak in the summer and minimum in the Northern Hemisphere winter [*Baughcum et al.*, 1996a]. *Baughcum et al* [1996a] noted a global increase in fuel usage of 4.6%/yr from 1976 to 1992 (similar to the rate of increase quoted by *Hofmann* [1991]). Figure 1 shows average daily fuel burnt by region in 1976 as a function of season, according to the inventories presented by *Baughcum et al.* [1996b]. The North American/continental U.S. regions had the highest regional fuel consumption. Air traffic in these regions consumed 3 times the fuel of traffic in the European region and nearly an order of magnitude more fuel than traffic in the North Atlantic and North Pacific regions. Figure 2 shows fuel burnt by altitude for the month of May 1976 from the same inventory. It is seen that most fuel is consumed at cruise altitudes, ranging between 9.5 and 12.5 km msl pressure altitude (altitude range corresponding to the range of flight-level atmospheric pressures in the standard atmosphere). Depending on the region and season, more or less of this altitude band will be above the tropopause. Data for later years follow the same pattern, with respect to fuel consumption both by region and by altitude, with a slow overall year-by-year increase in each region and at each altitude, ranging in magnitude from 2.8% to 7.4% annually [*Baughcum et al.*, 1996a,b].

Although aircraft engines have continually become more efficient and average fuel sulfur content has decreased, leading to lower number concentration and sizes of emitted particulates, the continued increase in air traffic will make aircraft emissions a continuing potential pollution concern. As flight corridors become more congested, it is likely that more flight time may be spent in the stratosphere, where particulate residence times will be much longer and the impact on photochemistry greater. This will be especially evident in heavily traveled en route flight corridors, such as over the North Atlantic and North Pacific.

In order to compare to the large body of upper tropospheric aerosol measurements recently acquired in selected regions, we examine below a unique set of in situ aerosol observations obtained along flight corridors during the 1970s. These observations spanned much of the Northern Hemisphere and will be used to examine the atmosphere of that era for impacts due to aircraft emissions, and to formulate a baseline against which to compare current observations.

2. GASP

From March 1975 through June 1979, the National Aeronautics and Space Administration (NASA) conducted the Global Atmospheric Sampling Program (GASP). The purpose of the program was to develop a global database of important atmospheric species using aircraft-mounted automated instrumentation. These observations were to be used to improve understanding of the background chemistry and physics of the upper troposphere/lower stratosphere, and to provide a basis for evaluating the perturbations to this background state caused by aircraft emissions [*Perkins and Gustafsson*, 1975].

Similar atmospheric research programs involving automated equipment on scheduled commercial aircraft have been undertaken more recently. The Measurement of Ozone and Water Vapor by Airbus In-Service Aircraft (MOZAIC) program was initi-

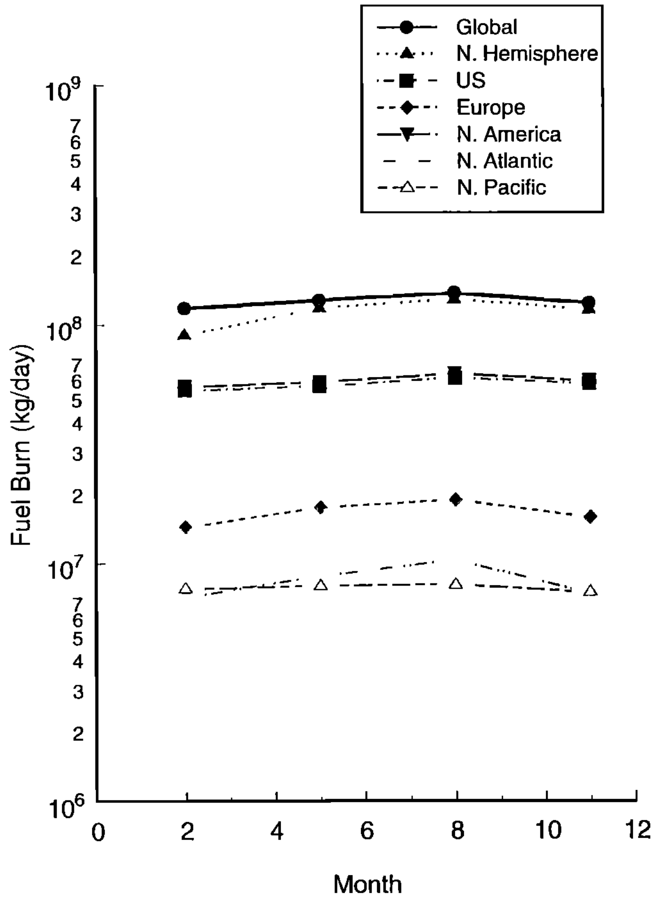


Figure 1. Commercial aircraft fleet fuel consumption during 1976 by region and season, re-plotted from *Baughcum et al* [1996b].

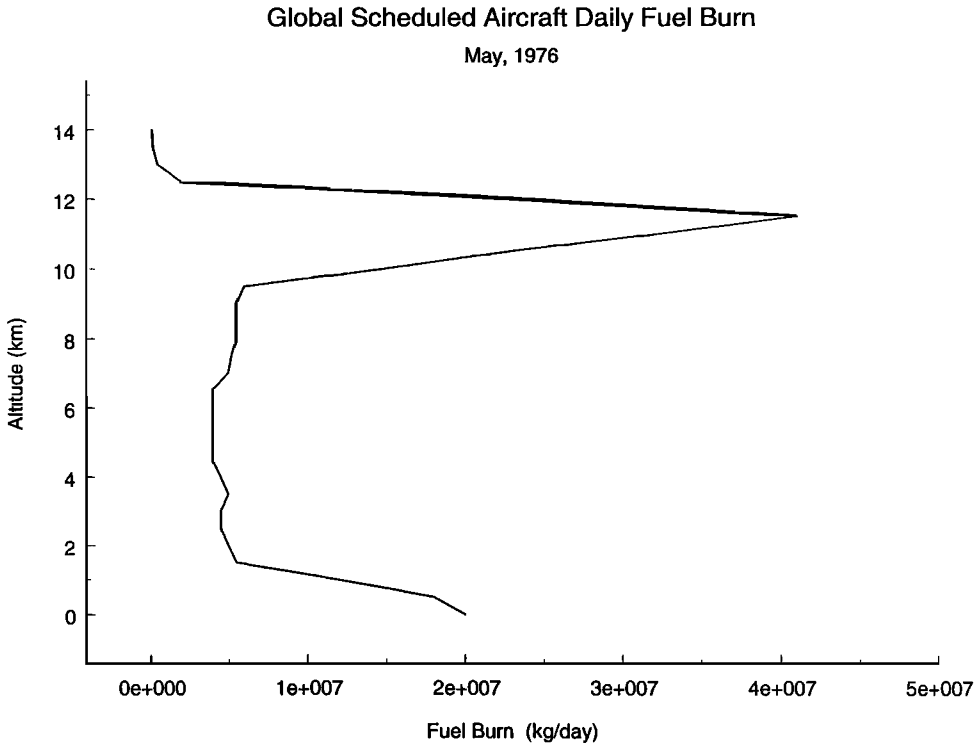


Figure 2. Global commercial aircraft fuel consumption by altitude, May 1976, from *Baughcum et al.* [1996b].

ated in 1993 and is still ongoing [*Marenco et al.*, 1998]. In addition to ozone and water vapor, instrumentation is being added to monitor other trace gases, such as carbon monoxide and oxides of nitrogen. The Nitrogen Oxides and Ozone Along Air Routes (NOXAR) program [*Brunner et al*, 1997] ran from May 1995 through May 1996, with automated instrumentation for measuring ozone, nitrogen monoxide, and nitrogen dioxide, on one scheduled Swissair B-747 aircraft. Neither of these programs included instrumentation for atmospheric particulates.

Descriptions of the GASP program are given by, for example, *Perkins and Gustafsson* [1975], *Nastrom et al.* [1982], and a series of data reports [e.g., *Holdeman et al.*, 1976]. Four Boeing 747s in routine commercial service were fitted with automated instruments to measure aerosol particle number concentration, size, and chemical composition and gaseous ozone and carbon monoxide concentrations. Other recorded parameters included wind speed and direction, pressure, temperature, and humidity. After the data were gathered and processed through quality control procedures, analyzed tropopause height information from the U.S. National Meteorological Center (NMC, now National Center for Environmental Prediction) operational hemispheric analyses, and computed solar elevation angle, were added during archival.

Owing to changing programmatic priorities and the premature termination of the program, analysis of the GASP data did not proceed as originally planned. Several preliminary analyses were published during or immediately after the field observation phase. Ozone data were used in several analyses. Selected portions of the O_3 data were used in studies of seasonal variations of ozone mixing ratio in the upper troposphere/lower stratosphere [*Nastrom*, 1979]. The complete data set was used to address the issue of exposure of flight crews to hazardous levels of O_3 in several reports, [e.g., *Jasperson et al.*, 1984]. The CO data are extensively discussed only in a contractor report [*Wu*, 1981], in which then novel aspects of the global distribution of CO in the upper troposphere were presented. Data from GASP optical particle counters have been used to indicate the presence or absence of cloud at flight level, and have served as a basis for compiling a global climatology of upper tropospheric cloudiness [e.g., *Jasperson et al.*, 1985]. Chemical analysis of particulate matter collected on filters from the limited effort in filter sampling is summarized by *Lezberg et al.* [1979].

Several reports appeared based on observations obtained on a special 2-day pole-to-pole around-the-world flight sequence organized during GASP. (See, for example, *Gauntner et al.* [1979], *Newell and Gauntner* [1979], *Pratt and Falconer* [1979], and *Hogan and Mohnen* [1979].) These reports focused on the distribution of O_3, CO, and aerosols along the flight track during this 2-day period.

Some analyses developed that were not anticipated by the original project planners. For instance, the GASP meteorological data served as the basis for several studies of atmospheric wave spectra [e.g., *Nastrom and Gage*, 1985; *Nastrom and Fritts*, 1992].

Despite original intentions, the data have yet to be critically applied to the problem of evaluating the effects of aircraft particulate emissions on the upper troposphere/lower stratosphere. The purpose of this paper is to reexamine the GASP data in light of current concerns over the impact of aircraft on the atmosphere We evaluate the extent to which portions of it can be used to establish a baseline for condensation nuclei (CN) (i.e., submicrometer aerosol particles [Pruppacher and Klett, 1997]) in the upper troposphere/lower stratosphere of the late 1970s, and the extent to which it can be used to discern the effects of aircraft on the distribution of CN in that era.

3. GASP Instrumentation

3.1 Data Acquisition

During GASP, data were gathered at altitudes greater than 6 km. A special cover protected the external sample inlet probes when the aircraft was below 6 km altitude to limit exposure of the intake lines to more heavily polluted lower tropospheric air. Data typically were recorded every 5 min. The data were, in most cases, observations obtained near the time of recording, and in some cases, an average of constituent concentrations over more or less of the preceding 5 min. At cruise velocities this resulted in observations at 75 km intervals along the aircraft route. Exceptions to this included periods when automated instrument calibrations were completed (calibration results but no observations recorded) and, later during GASP, when an aircraft encountered turbulence. In this latter case, data were in some cases recorded every second in an effort to gather more information about transport in the vicinity of jet streaks. On about 100 flights, observations were recorded at 1 s intervals for the entire flight, regardless of turbulence or location relative to jet streaks. Each aircraft operated for about 10 hours per day. This resulted in about 120 data or calibration points per aircraft per day from the normal recording protocol, and many thousands of points per day from flights with the 1 s observation interval.

All of the instruments were calibrated in the laboratory prior to installation. About every 2 weeks, data cassettes in the data acquisition systems and other expendables such as calibration gases, filters, etc., were replaced, and at this time, instrument calibration data were also taken. Instruments themselves were routinely swapped out for refurbishment and bench calibration. The interval for removal for refurbishment and bench calibration varied from a few weeks to many months, depending on the type of instrument and how often a given airliner passed through an airport where this could be done.

During archival of the observations, tropopause data were interpolated to the aircraft latitude and longitude from NMC analysis grid points. During the GASP program, NMC used two different methods to calculate the tropopause altitude [Holdeman *et al.* 1976]. The Flattery method, based on the lowest stable layer with a pressure $\leq$ 500 mbar and a lapse rate $\leq$ 2.5 K km^{-1}, was used from the start of the GASP program until December 17, 1975, when the Gustafsson method was implemented. This more complex method is based on potential temperature and the anticipated tropopause height for the season and the location. In the fourth GASP data report [*Holdeman et al.*, 1976], as well as later data reports, it is noted that the Gustafsson method generally produces a lower tropopause than the Flattery method and that this discrepancy increases as latitude decreases. The NMC went back to using the Flattery method in July 1977. No adjustments were made to the archived GASP tropopause data to adjust for this brief change in methodology. Since the CN data were all obtained after the Flattery method was reinstated, the definition of tropopause height used in the analyses below is consistent throughout the time span of the data.

3.2 GASP Condensation Nuclei Instruments

All of the instruments of the GASP package were derived from commercially available instruments that were altered to increase sensitivity and adapted for automated operation on board the aircraft. The GASP CN instrument was a modified Environment One CN counter. This instrument was designed by T.A. Rich, and derived from (but not identical to) the earlier General Electric (GE) CN counter by T.A. Rich and B. Vonnegut [*Skala*, 1963]. These counters began flying as part of GASP in October 1977.

Briehl [1974] and *Nyland* [1979] provided descriptions of the instrument and sample conditioning. It detects CN by producing rapid expansion of humidified sample air, resulting in formation of a cloud of water droplets. The CN concentration is determined by measuring extinction of white light in the resulting cloud. The commercial version of the instrument was modified for GASP for greater sensitivity at small concentrations and automatic operation and to operate with a pressurization system which brought the ambient air sample up to cabin pressure and temperature using filtered, particle-free air.

Tests on a prototype instrument showed that it detected concentrations from less than 10 cm^{-3} to 1000 cm^{-3} with a noise level of less than 5 cm^{-3}. *Briehl* [1974] showed that losses of involatile particles in the pressurization system were less than 20%, while *Nyland* [1979] showed average losses of 4.5%. Tests comparing the response of the GASP CN counter to that of a Pollak (expansion-type) CN counter for laboratory aerosols showed that the concentrations indicated by the GASP instrument were linearly proportional to those from the Pollak counter, but that relative response varied by ±25% depending on particle type [*Nyland*, 1979]. These results are consistent with tests performed using similar instruments by *Hogan et al.* [1981].

3.3 Analysis Procedures

The archived GASP data were obtained from the National Center for Atmospheric Research in Boulder, Colorado. These data were archived by aircraft. For each aircraft there are individual data files for each flight. *Holdeman et al.* [1976] provided a description of the data format. The CN data used in our analysis are an average over 16 s prior to the time the data are recorded. Recording occurs at 5 min intervals.

As we desired to identify a particular observation as stratospheric or tropospheric, only observations including ozone and tropopause height were used in the analyses reported below. This eliminated only a small fraction of the total data from the analysis, and gave us a second criterion in addition to the archived tropopause height for assigning an altitude as either tropospheric or stratospheric. *Wozniak* [1997] found that the archived NMC-analyzed tropopause heights were typically within 2 km of a distinct level at which ozone began increasing rapidly with height. Thus the definition of the boundary between troposphere and stratosphere at a particular location can differ by a distance of the order of 2 km, depending on the criteria used. There was no consistent bias between the two definitions of the tropospheric/stratospheric boundary. In the analysis described here, we use the archived NMC tropopause height to define this boundary.

The individual observations were arranged into seasonal groups according to seasons in the Northern Hemisphere: winter (December, January, February); spring (March, April, May); summer (June, July, August); and autumn (September, October, November). In addition, the data were divided into nine (sometimes overlapping) regions bounded by the latitude and longitude ranges indicated in Table 1.

Figure 3 shows regions defined in Table 1 plotted on a map of the Northern Hemisphere. Also plotted are the geographic locations of the 4788 CN observations in clear air incorporated into the analysis presented below.

Table 1. Regions and Ranges of GASP Analyses

Geographical Region	Latitude Range	Longitude Range
Northern Hemisphere*	0° – 80°N	180°W – 180°E
Continental United States*	25°N – 49°N	125°W – 70°W
Europe*	37°N – 70°N	10°W – 25°E
North America*	25°N – 70°N	125°W – 70°W
North Atlantic*	30°N – 70°N	70°W – 10°W
North Pacific*	30°N – 65°N	120°E – 125°W
Southeast Asia	0° – 30°N	60°E – 140°E
Western Pacific Ocean	0° – 30°N	130°E – 180°E
Hawaii	5°N – 35°N	172.4°W – 144.4°W

*Indicates regions that correspond to regions used in NASA's Scheduled Civil Aircraft Emission Inventories [*Baughcum et al*, 1996b].

3.4 Data Quality Evaluation

From studying selected individual flights in detail, it was found that relatively higher CN readings in the troposphere often occur in association with clouds, as indicated by the GASP optical particle counters. In many cases these clouds could be in a region of recent upward transport, which in addition to bringing moisture upward, also bring CN up from the surface boundary layer to the upper troposphere. It also is possible that evaporating cloud particles acted to humidify the sample air stream and affect the response of the CN counters or that cloud particles shattering on the inlet walls artificially enhanced CN counts [*Weber et al.*, 1998]. For these reasons, only cloud-free observations are included in statistics on which our CN analyses are based, with the goal that our results be representative of clear air volumes not recently influenced by relatively rapid vertical transport or high humidity.

An observation was determined to be cloud free if the optical particle counter included in the GASP instrument package indicated a similar cumulative concentration of particles greater than 1.4 µm and 3.0 µm. Detailed study of data from individual flights, along with corroborating archived standard meteorological data, showed that in the presence of cloudiness the optical particle counter registered similar cumulative counts in these two archived size categories. This criterion was adopted in preference to other cloudiness parameters archived in the GASP data based on a simple threshold count in the >3.0 µm optical particle counter channel. These other cloudiness parameters were based on count statistics covering the entire 5 min between recorded data points, while we required an indication of cloudiness sensitive specifically to clouds near the time of the observation because the CN observations represented an average over only 16 s prior to the recording time.

There were over 20,000 archived individual observations containing CN, ozone, and tropopause height information during the 21 months of GASP in which CN counters were flown. Selecting for observation times also containing optical particle counter data, and then for cloud-free observations according to the above criteria, left 4788 observations for the analysis described below.

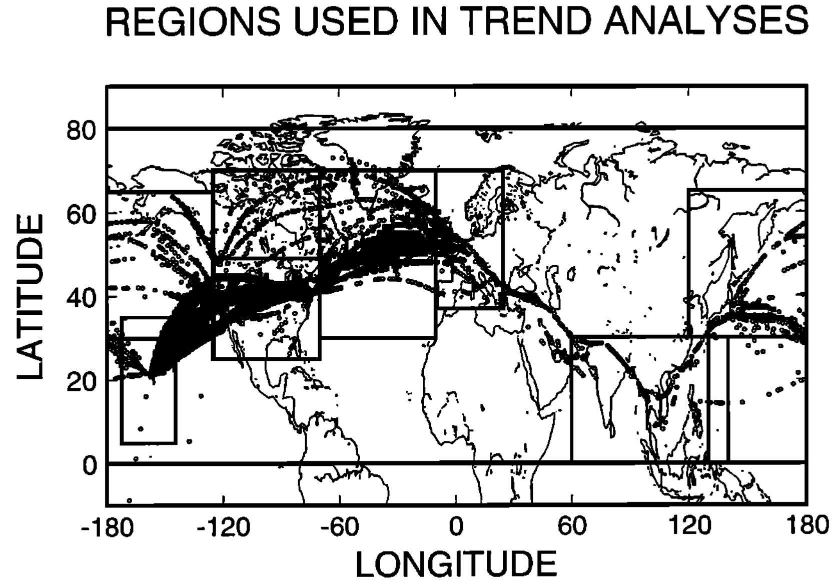


Figure 3. Regions used by *Baughcum et al.* [1996b], also adopted as the basis for the analyses performed in this study, with the addition of the areas centered on Hawaii and southeast Asia. Superimposed are locations of the 4788 CN observations utilized in the analysis.

4. Comparison to Other Observations of Condensation Nuclei

4.1 University of Wyoming Aerosol Measurements in the Upper Troposphere and Lower Stratosphere

For more than two decades there has been a continuous series of vertical soundings through the troposphere and stratosphere monitoring aerosols and trace gases conducted by researchers at the University of Wyoming from a site near their campus. (See, for example, *Hofmann* [1993] and *Hofmann et al* 1998].) One instrument often carried on these flights is a CN counter. In the Wyoming CN counters, ambient particles drawn into the instrument are first passed through a thermal diffusion chamber with ethylene glycol as a working fluid. A temperature gradient in the chamber results in large glycol supersaturations, causing condensation of glycol on all particles present followed by growth to a size where they will scatter enough light to be detected individually as they pass through a light beam [*Rosen and Hofmann*, 1977]. These condensation nucleus counters, although of a different design than the GASP condensation nucleus counters, are sensitive to particles of similar composition and in approximately the same size range. *Cadle et al.* [1975] showed that CN counters of design similar to the GASP counters produced concentration readings within 30% of those from instruments of the type flown on the Wyoming balloons, when sampling a laboratory-generated sulfuric acid aerosol.

It should be noted that the Wyoming balloon-borne CN counter is optimized for CN concentrations typical of the stratosphere. At altitudes below ~15 km a dilution system is employed in order to avoid exceeding the counting capability of the instrument [*Rosen and Hofmann*, 1977]. This introduced an additional uncertainty into the concentration estimate not present in the *Cadle et al.* [1975] experiments [*Hofmann*, 1993]. *Rosen and Hofmann* [1977] estimated a total uncertainty in absolute concentration of as much as a factor of 2 for this instrument, but this does not include uncertainty in the dilution.

A search was performed for GASP data obtained near Laramie, Wyoming, on days when there were balloon-borne instruments monitoring CN over this location. Several cases were found when a GASP-instrumented aircraft passed within a few degrees of latitude/longitude of Laramie on the day of a balloon launch. Given typical wind speeds at flight levels, even a few hours' separation can mean separation between measured air parcels of several hundred kilometers in a Lagrangian sense. However, based on examination of data from numerous individual GASP flights, upper tropospheric aerosol particle concentrations typically do not vary significantly in the horizontal at a given altitude over distances of many hundreds of kilometers, provided this region does not span an obvious meteorological feature such as a sloping tropopause or a jet stream axis. By carefully examining the circumstances when one or more GASP flights crossed in the vicinity of a balloon launch, we found four cases with good

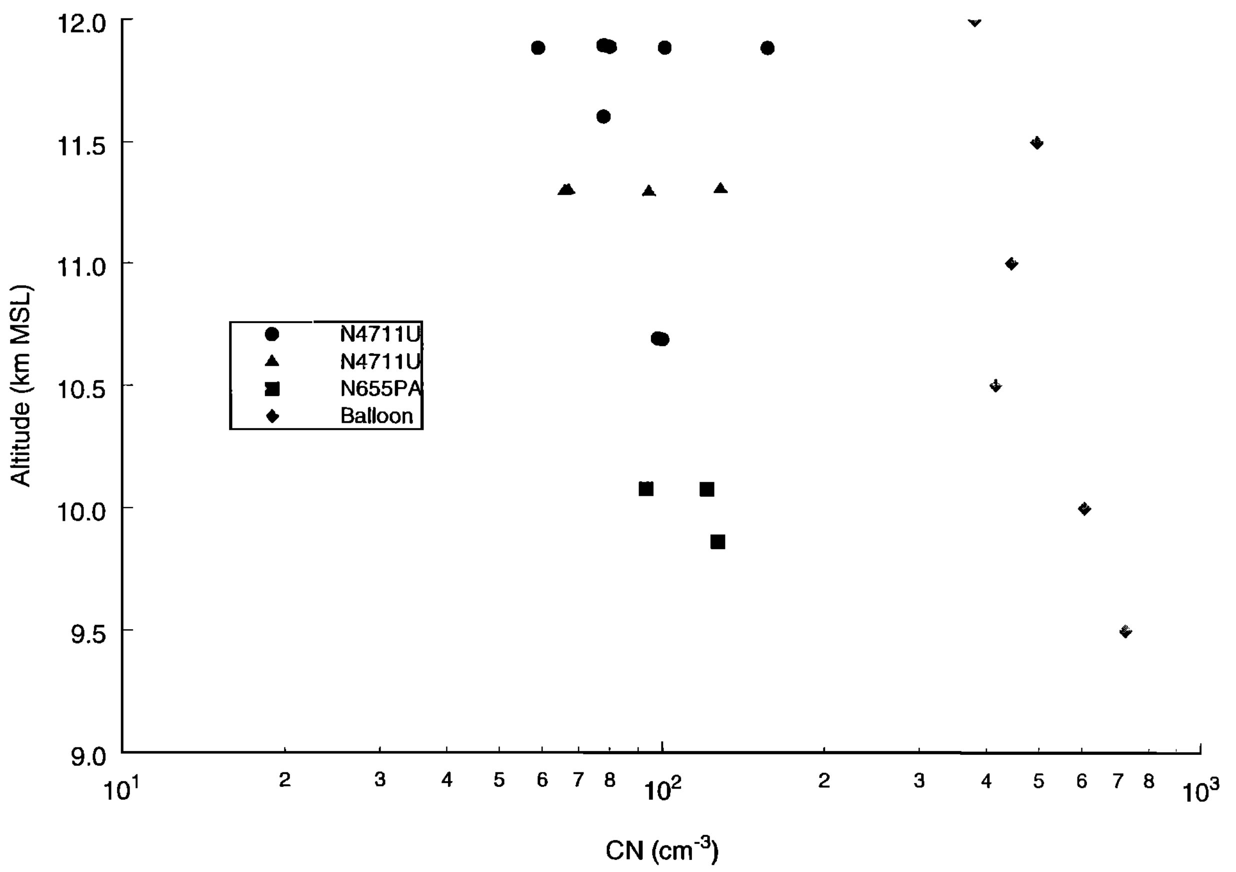


Figure 4. Vertical profiles of CN concentrations (number per ambient cubic centimeter) near Laramie, Wyoming, on June 13, 1978. Observations are presented for three passes by two different GASP aircraft, one passing Laramie twice during the day, and from a dawn balloon launch (diamond symbols).

reason to believe the aerosol sampled by the balloon-borne instruments and the GASP instruments was similar. For two of these cases, CN data are available both from the balloon instrument package and the aircraft instrument packages.

Figure 4 shows a vertical profile of CN concentrations derived from two different GASP aircraft, one passing near Laramie twice during the day, on June 13, 1978, and from a dawn balloon launch on the same day. (Note that data plotted are concentrations per unit volume of ambient air in contrast to later figures in which concentrations per unit mass of air are plotted.) The balloon instrument data are averaged over 500 m altitude intervals (~100 s of rise) corresponding to a sample volume of ~1.66 L. Each GASP reading is an average over 16 s (~ 4 km) of horizontal flight corresponding to a sample volume of ~1.33 L. Thus each point is based on counts of 10^5 or more particles, in the case of the Wyoming instrument, and extinction of light by a cloud containing ~100 cm^{-3} droplets in the GASP CN counter. These are conditions in which both instruments can yield statistically stable readings [*Hogan,* 1966]. There are multiple GASP CN readings at each level, corresponding to the period when an aircraft was within 5° latitude/longitude of Laramie on one of its passes near the city at that level (up to eight readings were obtained within 5° of Laramie on each pass depending on the aircraft trajectory). The two aircraft passed near Laramie from 6 hours before until 12 hours after the balloon launch. On one pass, one aircraft changed altitude near Laramie, yielding points over a range of altitudes. Meteorologically, there was a broad upper level ridge over the region on June 13, 1978. Surface observations indicated scattered isolated convective clouds later in the day; upper level winds were light and the ridge moved only slightly during the day, giving a reasonable expectation that the balloon and aircraft instruments sampled aerosol with similar properties.

It can be seen in Figure 4 that within the GASP observations on this day, there are increasing numbers of observations and increasing variability as altitude increases, over the altitude range sampled. The balloon CN concentrations are at least a factor of 3 and up to a factor of 6 higher than the median GASP CN concentrations at any level.

Another comparison on August 4, 1978, involving one of the same aircraft involved in the intercomparison shown in Figure 4, also indicated a factor of 5 to 6 higher CN concentration measured by the balloon instrument compared to the GASP instrument. The aircraft passed near Laramie ~6 hours after the balloon launch on this day. The bias between the GASP and Wyoming CN counters is consistent for two different GASP counters, and for one of those counters on two days separated by almost 2 months.

These limited intercomparisons suggest a distinct bias between the GASP CN counters and the Wyoming CN counters, with the GASP counters reporting concentrations significantly lower than the Wyoming counters. This bias is consistent with the difference between the center of the frequency distributions of GASP CN concentrations shown by *Wozniak* [1977], and of Wyoming CN observations shown by *Hofmann* [1993] and *Hofmann et al* [1998]. Typical GASP CN concentrations are in the range from several tens to a couple of hundreds per cubic centimeter, while typical upper tropospheric/lower stratospheric Wyoming CN concentrations range from a few hundreds to many hundreds per cubic centimeter. This is despite the fact that the intercomparisons reported by *Cadle et al.* [1975] would lead one to expect reasonable agreement. Inlet losses, as discussed by *Huebert et al* [1990], were not evaluated by GASP researchers. Such losses are estimated to be no larger than of the order of 10% for submicrometer aerosol particles and cannot explain the discrepancy between GASP and Wyoming CN observations. Concentrations are high enough that statistical counting noise [*Hogan,* 1966] cannot explain the disparity, either. In order to find plausible physical reasons for this discrepancy, a survey of other upper tropospheric CN measurements was conducted.

4.2 Comparison to Other Upper Tropospheric Aerosol Measurements

Other sources of in situ upper tropospheric CN data from the late 1970s include *Hogan and Mohnen* [1979], *Hogan* [1981], and *Hogan* [1986]. *Hogan and Mohnen* [1979] underflew a GASP aircraft on a second aircraft in the middle troposphere. Their aircraft was equipped with a modified Pollak CN counter (an expansion counter condensing water, of the same design as the counters used to calibrate the GASP CN counters in the laboratory [see *Hogan,* 1986]). The sample stream on this aircraft came from the compressor bleed air from the turboprop engines. Observations were obtained in clear air. The concentrations reported from the two aircraft near the same location but at different altitudes were in general agreement, showing a gradual decrease with height similar to that observed in typical Wyoming vertical CN profiles. Concentration magnitudes reported from both aircraft, though, were similar to the GASP CN concentrations shown in Figure 4, again much lower than typical concentration values from the Wyoming CN instrument. *Hogan* [1981] and *Hogan* [1986] report middle and upper tropospheric CN concentrations (also derived with Pollak-type counters) in remote regions of the clear middle troposphere in the range from many tens to several hundreds per cubic centimeter, generally consistent with the range of concentrations within the GASP data set (described more fully, below).

Beginning in the late 1970s and continuing through the 1980s the dominant technology in airborne CN counters changed from expansion-type cloud chamber counters condensing water to thermal diffusion chamber counters condensing organic vapors, like those used in the Wyoming balloon sounding program and those manufactured by Thermal Systems, Incorporated (TSI) of Minneapolis. The TSI CN counters are diffusion chamber systems similar in principle to the Wyoming counters, but usually butanol is used as a condensing fluid in the TSI counters in contrast to the ethylene glycol used in the Wyoming counters. A survey of selected recent CN measurements in the middle and upper troposphere obtained with these diffusion chamber counters [e.g., *Clarke,* 1993; *Bodhaine,* 1996; *Schroder and Strom,* 1997; *Ridley et al.,* 1997; *Anderson et al,* 1998a,b; *Paladino et al.,* 1998; *Miake-Lye et al.,* 1998; *Hagen et al.,* 1998] shows that in situations where the sampling includes only those particles that are predominantly non volatile and larger than ~0.02 μm diameter, the diffusion chamber counters yield concentrations in the same range as those observed during GASP. When sampling with diffusion chamber instruments like the TSI 3760 CN counter is done at ambient pressure and with only modest increases in temperature as the sample flows through the sampling instrument, reported CN concentrations are similar to those obtained with the Rosen CN counter on the Wyoming balloon launches, and a factor of 3 to 6 greater than typical GASP CN concentrations.

4.3 Summary of Data Quality Issues

We suggest that disagreement between the contemporaneous GASP and Wyoming CN measurements in 1978 may be due to loss of some fraction of the ambient CN during compression in the GASP system, or to differences in response to volatile or semivolatile CN between water-based expansion-cooling counters, and organic-fluid-based diffusion chamber instruments. The losses of particles greater than 0.02 μm diameter in the GASP system must have been due to processes other than diffusive losses to walls of the sampling system; these diffusive losses were shown to have been small by *Nyland* [1979]. Periodic laboratory calibrations of the GASP CN counters during GASP, comparisons between two operational GASP counters sampling in the same region on the same day, and comparisons of the same GASP instrument to the Wyoming instrument on two separate days suggest that the GASP CN data are internally consistent and that relative variations between regions and seasons, and between stratosphere and troposphere, as seen in the GASP data set, are reliable.

5. Results

Figure 5 shows the vertical variation of CN number mixing ratio for selected regions. Data were aggregated in 1 km altitude blocks relative to the tropopause height, by region and by season. Altitude relative to the tropopause was chosen for this presentation, as the distribution of CN mixing ratio in the vertical is observed to be nearly constant or slightly decreasing with altitude in the troposphere, and more strongly decreasing with altitude in the

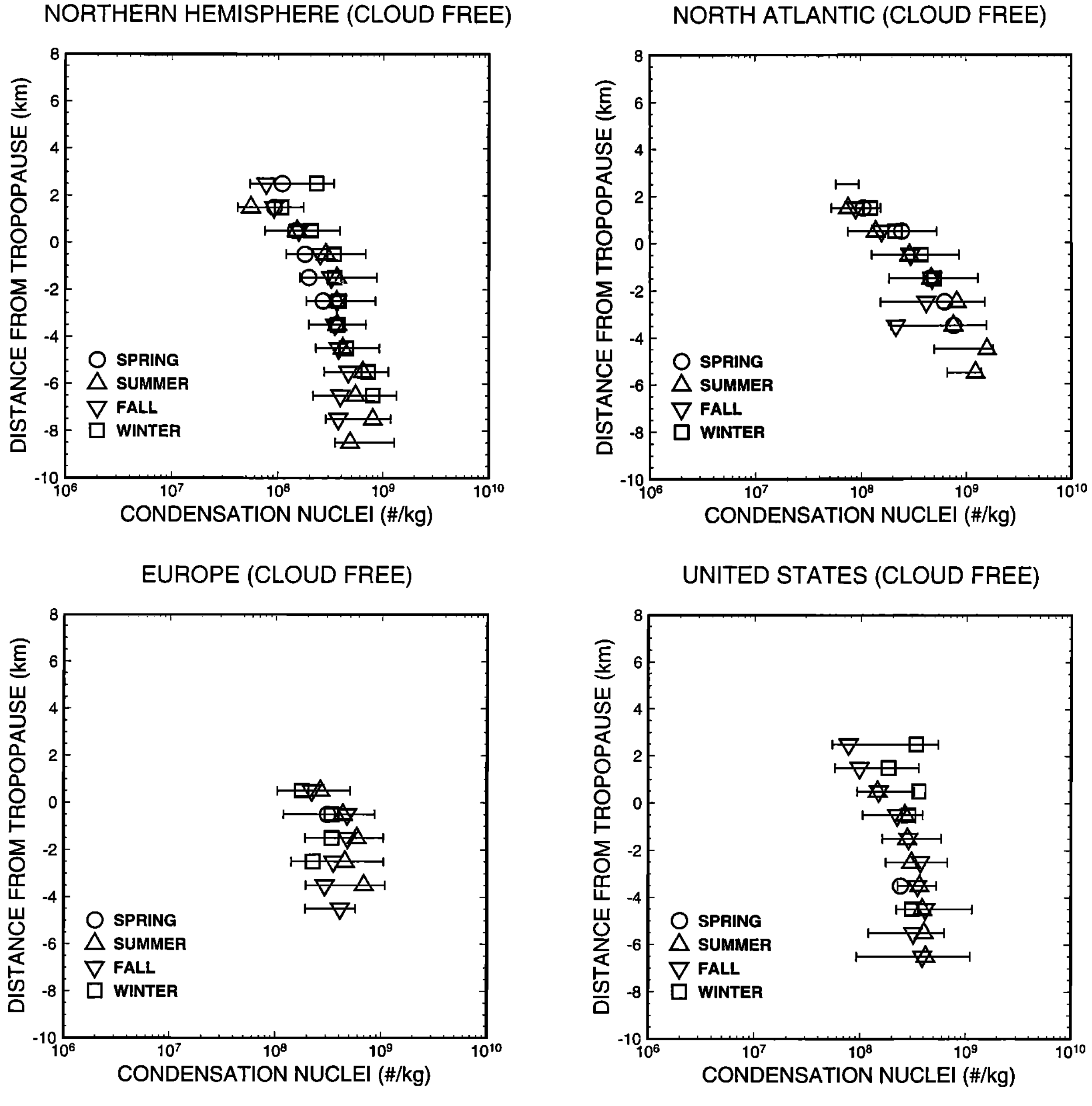


Figure 5. Vertical variation of GASP CN mixing ratios (number per kilogram of air) by region and season, relative to the tropopause. Regions are defined in Figure 3. Central symbols represent seasonal median mixing ratios, and whiskers represent 10th and 90th percentiles of all data at the indicated level.

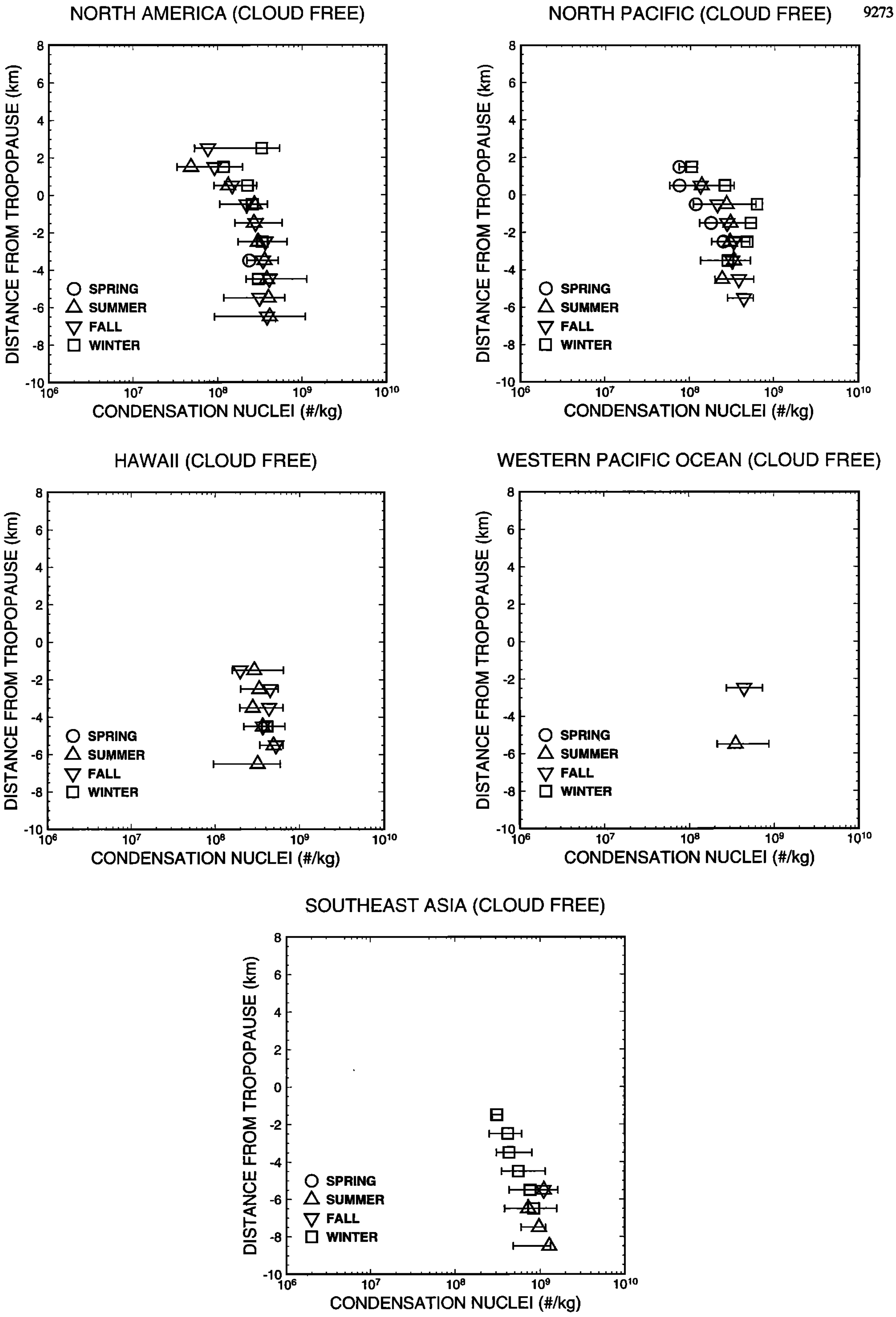


Figure 5. (continued)

stratosphere, in all regions and seasons. As the tropopause height itself varies with the state of the atmosphere and the season, the pattern of variation of mixing ratios, versus altitude above the ground, is less consistent (not shown). At each altitude relative to the tropopause, seasonal medians are plotted when more than five observations are available in that altitude block in a season. Horizontal bars within each altitude block span the range from the 10th to the 90th percentile of all of the observations at that level. The number of points used to determine the median at an altitude within a region and a season varied from 5 to hundreds, as shown in Table 2. Relatively smaller numbers of observations were available in the western Pacific, southeast Asian, and European regions. From tens to hundreds of observations at altitudes near the tropopause were often available within a season for the remaining regions. Mixing ratio, the ratio of the number of particles to the mass of a unit volume of air in which they are contained, is used in the analysis in order to eliminate changes with altitude in analyzed CN abundance due to expansion and compression of air parcels during vertical mixing. (At the 9 km msl level in the International Civil Aviation Organization (ICAO) standard atmosphere, corresponding to a pressure of 307 hPa, a mixing ratio of 10^8 CN kg^{-1} corresponds to a concentration of 47 CN cm^{-3}.) Typical tropospheric clear air mixing ratios fall in the range of 10^8 to 10^9 CN kg^{-1} in the GASP data set. Within the upper troposphere, the medians and ranges of CN mixing ratios vary little with height, within a given region and season. Near the tropopause the scatter of the points generally diminishes and the median mixing ratios begin to decrease with increasing altitude. Concentrations in the lowest 2 km of the stratosphere are generally slightly less than 10^8 CN kg^{-1}. This vertical distribution pattern suggests that CN are predominantly generated at the surface or in the troposphere and are relatively well mixed vertically through the clear troposphere, and that the troposphere is generally a source of CN for the stratosphere. Highest tropospheric concentrations are found in the southeast Asia, North Atlantic, and Europe regions. No clear variation with region or season is evident in the available stratospheric data.

Regions nearer the tropics (southeast Asia, western Pacific Ocean, and Hawaii) have a less distinct transition to lower values that occurs further below the tropopause than in the other regions. A similar discrepancy between a chemically defined tropopause and a thermodynamically defined one near Samoa has been described by *Folkins et al.* [1999]. The tropical tropopause is determined by the deepest convection globally (typically in the tropical western Pacific), and in other tropical regions, convection often does not reach this tropopause height. It might be inferred that the vertical distribution of CN is strongly affected by the vertical extent of local convective mixing in a region. There are few stratospheric data points in tropical regions, and the data are not sufficient to define typical lower stratospheric mixing ratios in the tropics.

The same medians are plotted in Figure 6, for each region, for each of the four seasons. The points representing the medians from each region at a given level are offset slightly in the vertical to make them more readily discernible. The bars through each point represent the range from the 10th to the 90th percentile in the distribution of mixing ratios in that region and season at that level. It can be seen that in the spring, at tropospheric levels within a few kilometers of the tropopause, the range of observed CN mixing ratios from the North Atlantic and Europe tends to be larger than for other regions, and generally displaced to higher values than for other regions or the Northern Hemisphere as a whole. These same regions tend to have higher values in summer and fall, as does southeast Asia. In winter the ranges at different levels from the different regions tend to overlap more closely.

Because comparisons of this sort can be difficult to present when data are plotted on a log scale, the vertical variations of median CN mixing ratios in the different regions are compared in Figure 7 as ratios to the overall Northern Hemisphere medians at the same altitude using a linear scale. Data are plotted where there are five or more regional readings at a given altitude in a given season. Table 2 shows the number of observations available at each level in each region and season. The number of data points from which a median was determined at a given altitude in a given season tended to be small, between five and ten, in the western Pacific, southeast Asian, and European regions. From tens to hundreds of observations at altitudes near the tropopause were often available in a given altitude block within a season for the remaining regions Relatively fewer observations were available in most regions as distances from the tropopause increased. The spring season in general had the smallest number of observations.

Most of the Northern Hemisphere observations are from the North American region, so there is little distinction in most of the seasons between the medians of the Northern Hemisphere, North American, and United States groupings. One exception is the relatively higher CN mixing ratios in the winter season over the United States, near and above the tropopause, compared to the overall Northern Hemisphere medians. The largest enhancements observed for any region in our study are the enhancements from 1 to 4.5 km below the tropopause over the North Atlantic in fall and winter.

In almost all seasons where sufficient cloud-free data are available, and at altitudes just below the tropopause, Europe and North Atlantic medians exceed the Northern Hemisphere medians. An exception to this pattern is Europe in the winter, when medians were similar to the overall Northern Hemisphere medians.

Regions over the Pacific are generally cleaner than the Northern Hemisphere as a whole. Exceptions to this pattern are mixing ratios somewhat exceeding the Northern Hemisphere medians just below the tropopause in winter in the northern Pacific regions, and well below the tropopause in the southeast Asia region in summer.

6. Discussion

Regional variations of CN at flight altitudes do not correlate closely with regional patterns of aircraft fuel consumption shown in Figure 1. The North Atlantic region, for instance, has generally higher median CN mixing ratios than the North America region at altitudes just below the tropopause, but lower regional fuel consumption.

The vertical distribution of CN does not correlate well with vertical distribution of fuel burn, either. Figure 5 suggests that CN are well mixed vertically in the cloud-free troposphere, with median number mixing ratios varying little with altitude in the middle and upper troposphere, and in some seasons and regions even beginning to decline with increasing altitude in layers a few kilometers below the tropopause. In contrast, aircraft fuel consumption is predominantly in the upper troposphere and lower stratosphere, within a few kilometers of the tropopause.

We discuss below some hypotheses for this lack of correlation. These include horizontal transport and mixing with the prevailing

Table 2. Number of Observations for Each Region, Season, and Altitude Level

	Northern Hemisphere				North America				United States				North Atlantic			
Level	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi
4				1												
3			3	1			2				2					
2	14	1	19	20		1	16	12			14	12			3	2
1	12	45	77	48		28	48	23		2	37	18	6		22	11
0	54	191	214	107	1	54	128	27	1	12	113	22	18		57	42
-1	37	323	313	111		39	189	27		17	180	22	3		87	33
-2	42	397	378	85		45	203	2		39	201	2	14		92	16
-3	17	276	365	55	2	26	166	6	2	25	166	3	6		34	4
-4	13	242	296	47	5	66	73	4	5	65	73	4	8		8	3
-5	2	154	277	51	2	30	36	6	2	30	36	6			3	2
-6	3	86	84	34		13	13	3		13	13	3				1
-7		73	16	12		5	10			5	10					
-8		11	8	4		1	3			1	3					

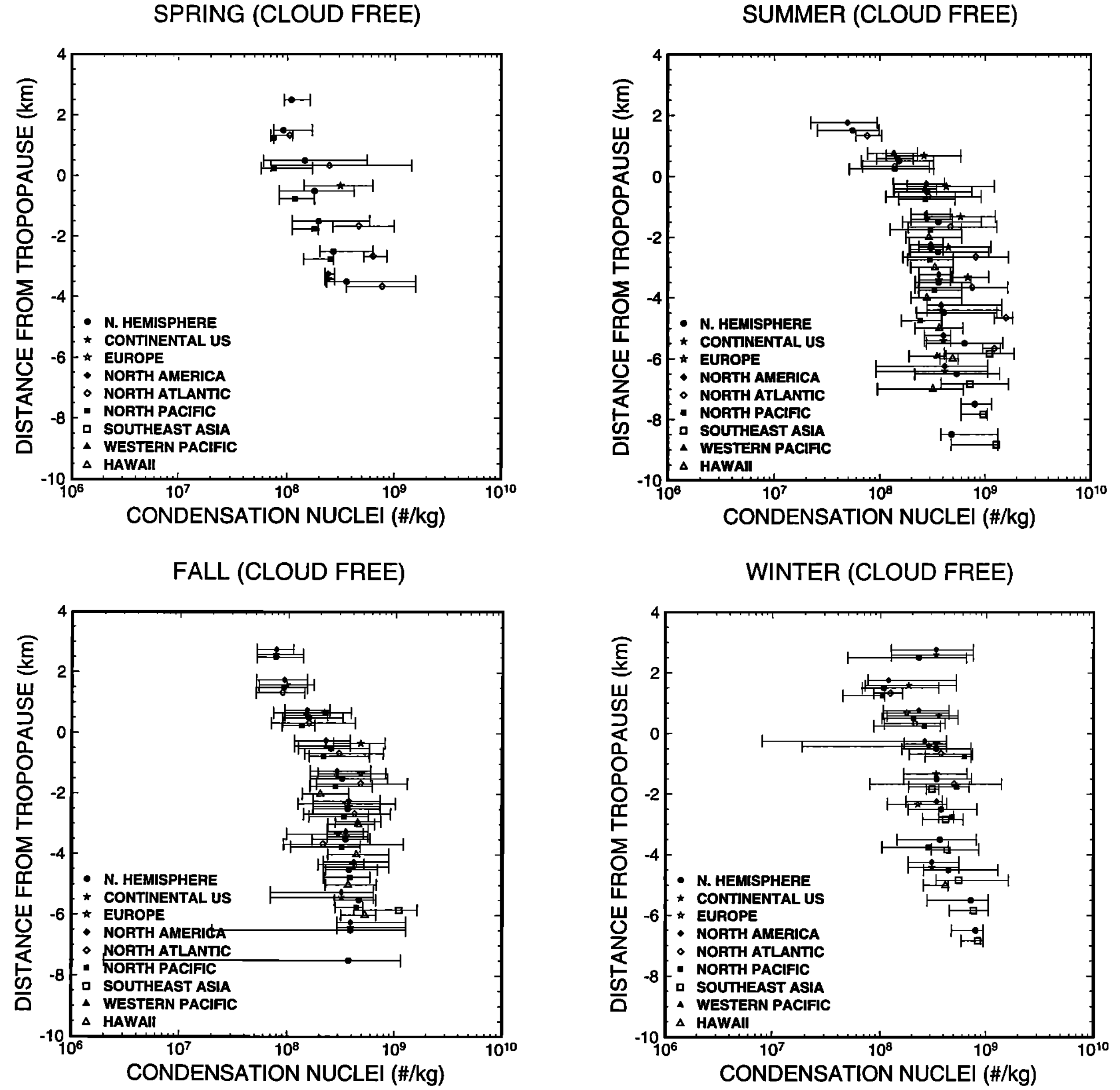
Europe				North Pacific				Hawaii				Western Pacific				Southeast Asia			
Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi
			1				4												
1			4	5	1	3	5												
3	23	8	13	21	49	5	8												
14	33	12	10	20	120	19	15		3					1	1			1	4
3	31	16	9	25	139	33	17		19	11	4		1	7				2	6
	22	15	10	9	69	90	8		56	24	1		2					1	10
	5	5	2		45	98	15		80	60	4		2	2	1			4	11
		5	1		9	87			55	93	12		4		2		4	4	14
1	1	4	1		4	16			9	24	3		9	2	2	2	9	7	22
	1								12	2	2		2		1		30	3	7
										3	1						6		3

Sp denotes spring, Su denotes summer, Fa denotes fall, and Wi denotes winter.

atmospheric circulation patterns as the CN mix vertically, the fact that globally the most important source of CN is fossil fuel combustion, and fuel combustion by aircraft represents only ~2% of total fossil fuel combustion, and finally, the possibility that the GASP CN counters were not sensitive to some component of the aircraft-produced CN.

Figures 6 and 7 suggest some seasonal variability in the ratios of regional medians to overall Northern Hemispheric medians. While we have not been able to do a comprehensive meteorological analysis for the complete period over which the GASP CN observations were obtained, it seems probable that variations in hemispheric atmospheric flow regimes, as well as low sample numbers in some altitude blocks within some regions, are important contributors to this variability.

Ferry et al. [1999] described aerosol sampling in a major flight corridor over the North Atlantic and showed an example in which aerosol particle concentrations are higher and the particles more volatile in an altitude range in which there is heavy air traffic, compared to lower altitudes. They show that 73% of the particles in the layer containing aircraft emissions were volatile at 120°C. While the GASP sampling system pressurizes the sample only to cabin pressure (roughly 850 hPa) and temperature (roughly 20°C), it is possible some loss of volatile particles occurs in the sampling system. We suggest that such a loss may explain the disparity between the Wyoming and GASP CN measurements in the same air mass. If this is true, then perhaps the vertical profiles shown in Figure 5 represent only relatively involatile particles and a set of instruments sensitive to smaller and


SPRING (CLOUD FREE)

SUMMER (CLOUD FREE)

FALL (CLOUD FREE)

WINTER (CLOUD FREE)

Figure 6. As in Figure 5, but with medians for each region shown as a function of altitude for a given season. Symbols at each altitude block are offset slightly to help distinguish each from the others. Bars through symbols span the 10th to 90th percentiles of the range of readings for that region and altitude. Data are presented for the seasons (a) spring, (b) summer, (c) fall, and (d) winter.

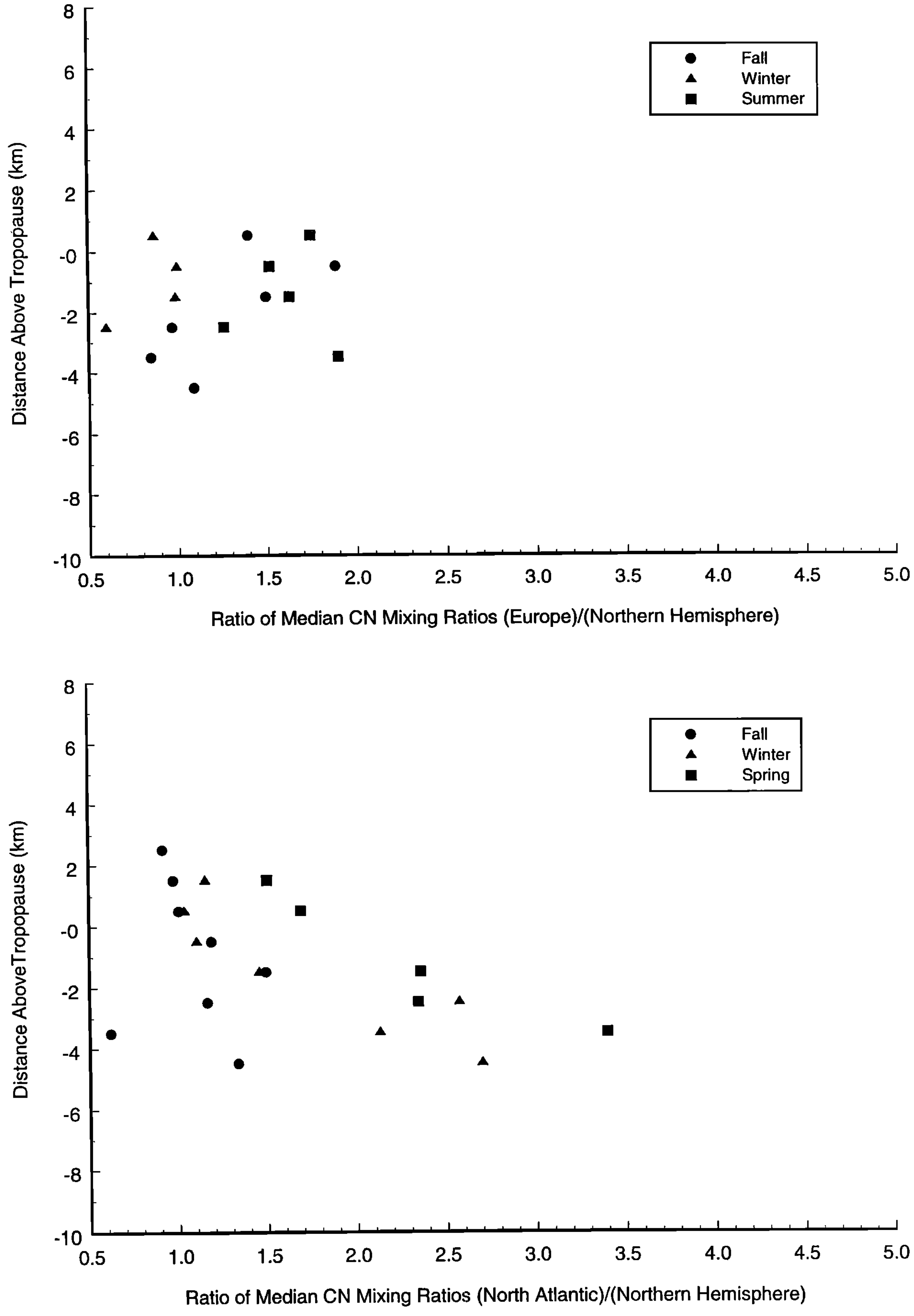
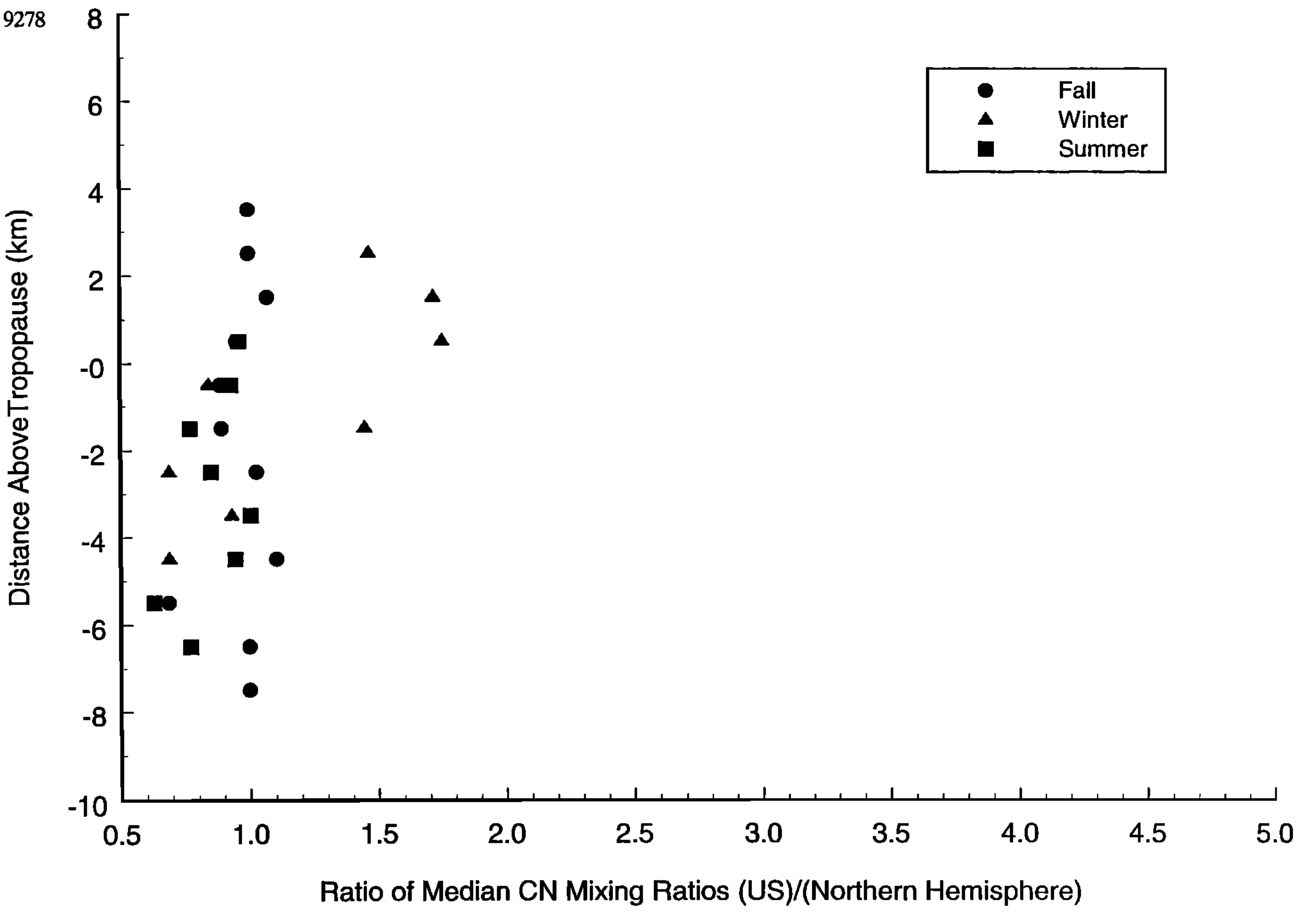


Figure 7. Ratios of regional median CN mixing ratios to Northern Hemisphere median CN mixing ratios at the same altitude, as a function of altitude relative to the tropopause, for each season. No data are plotted for seasons and levels when regions contain fewer than five observations at that level within the season.



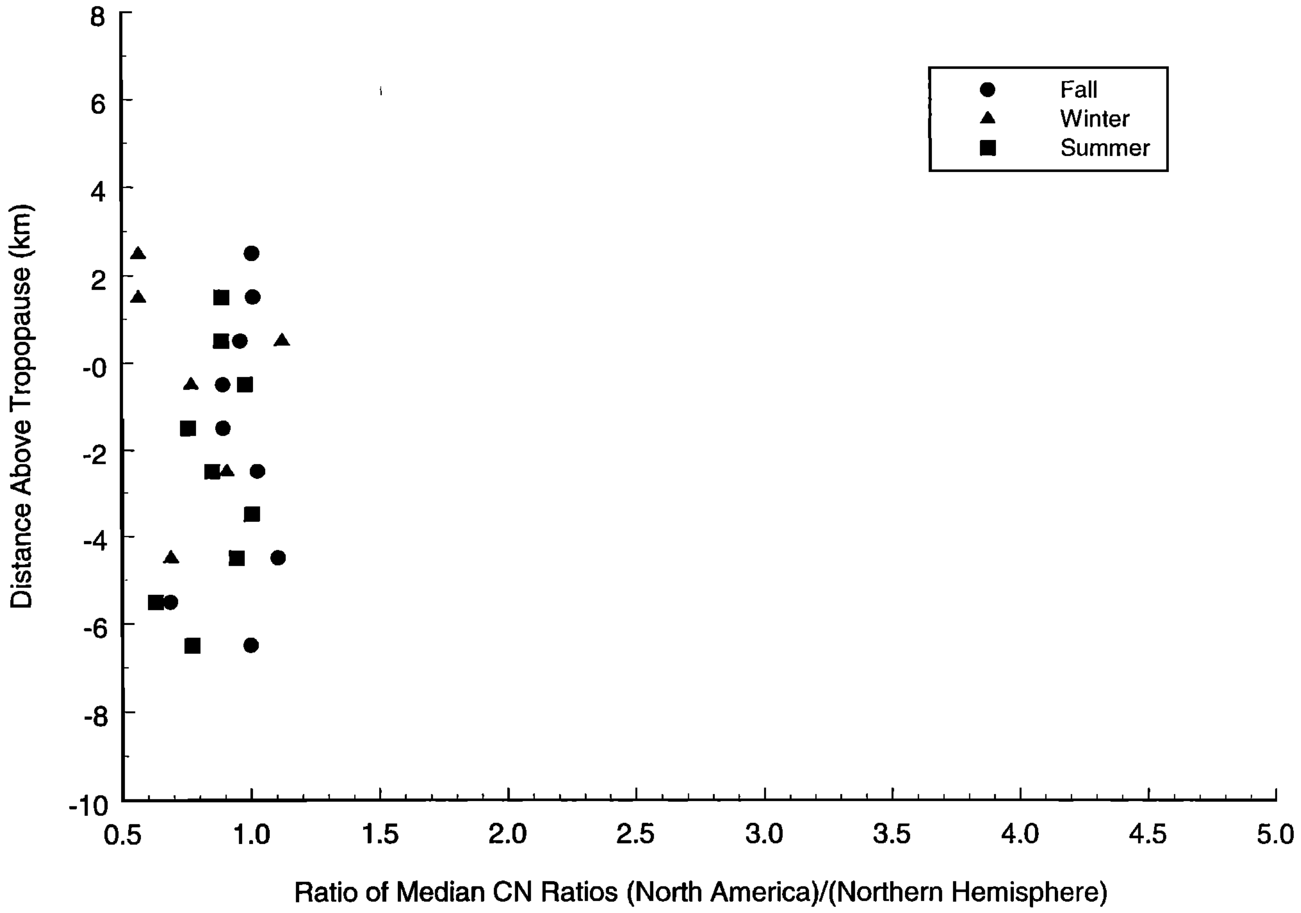
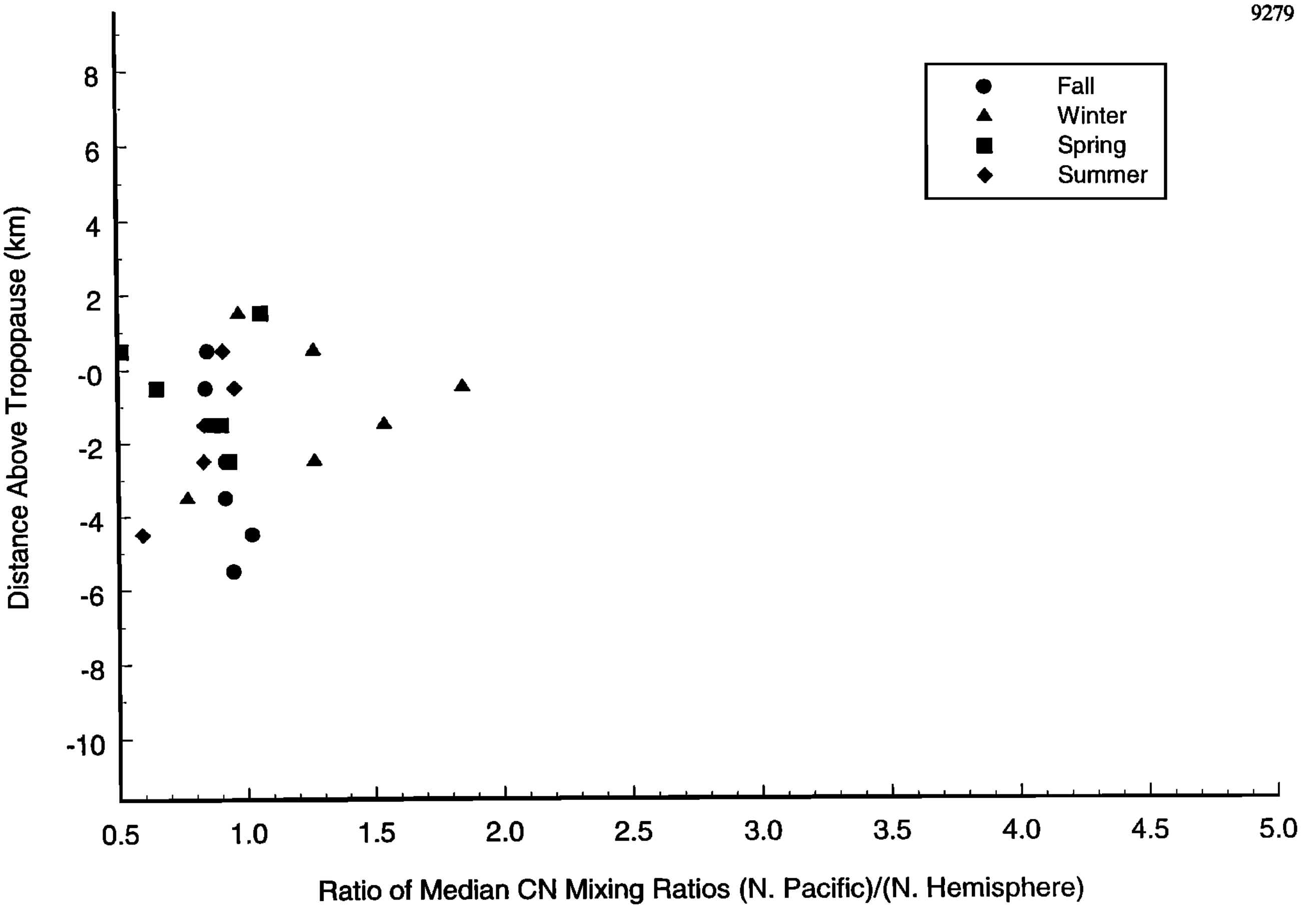


Figure 7. (continued)



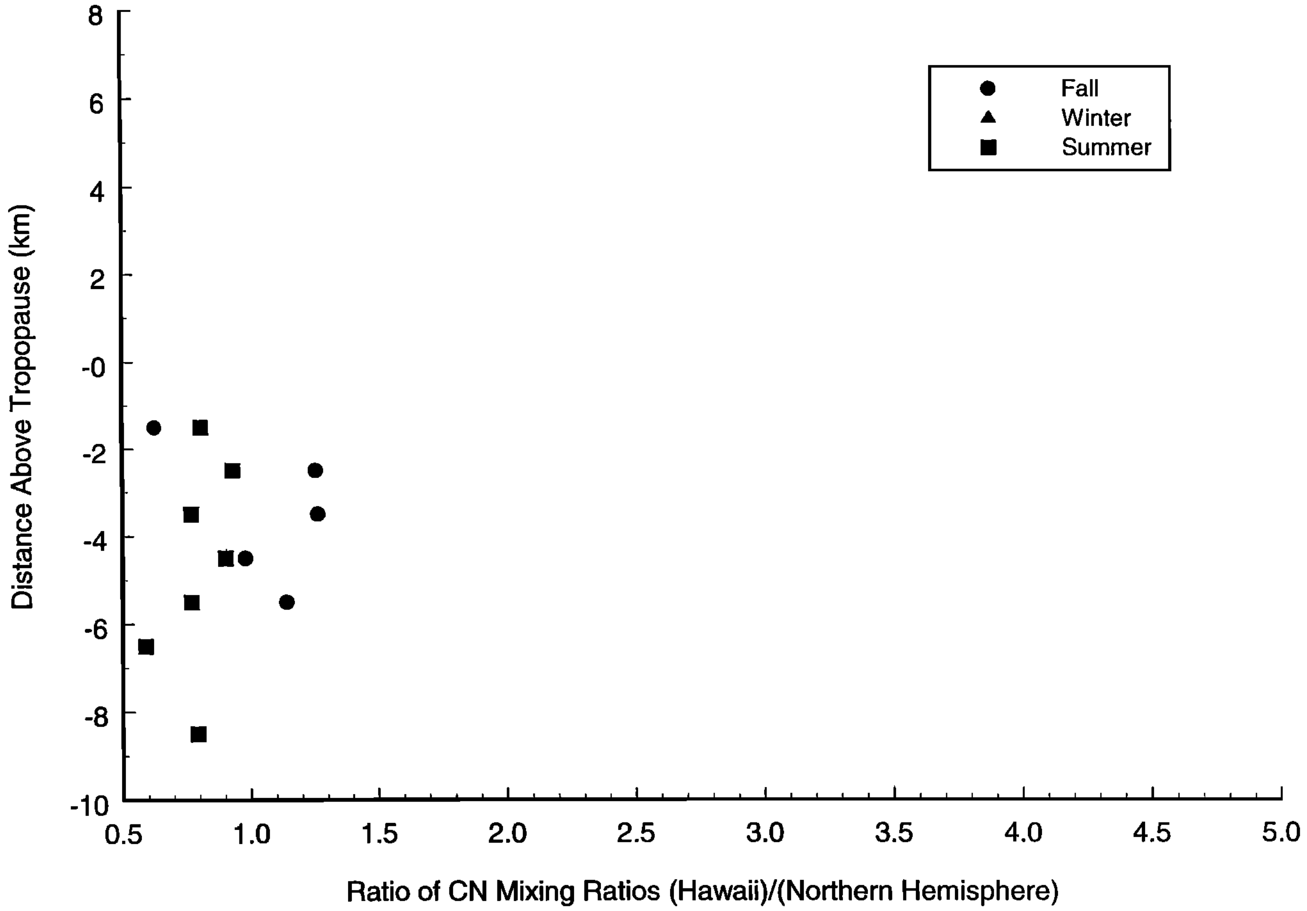
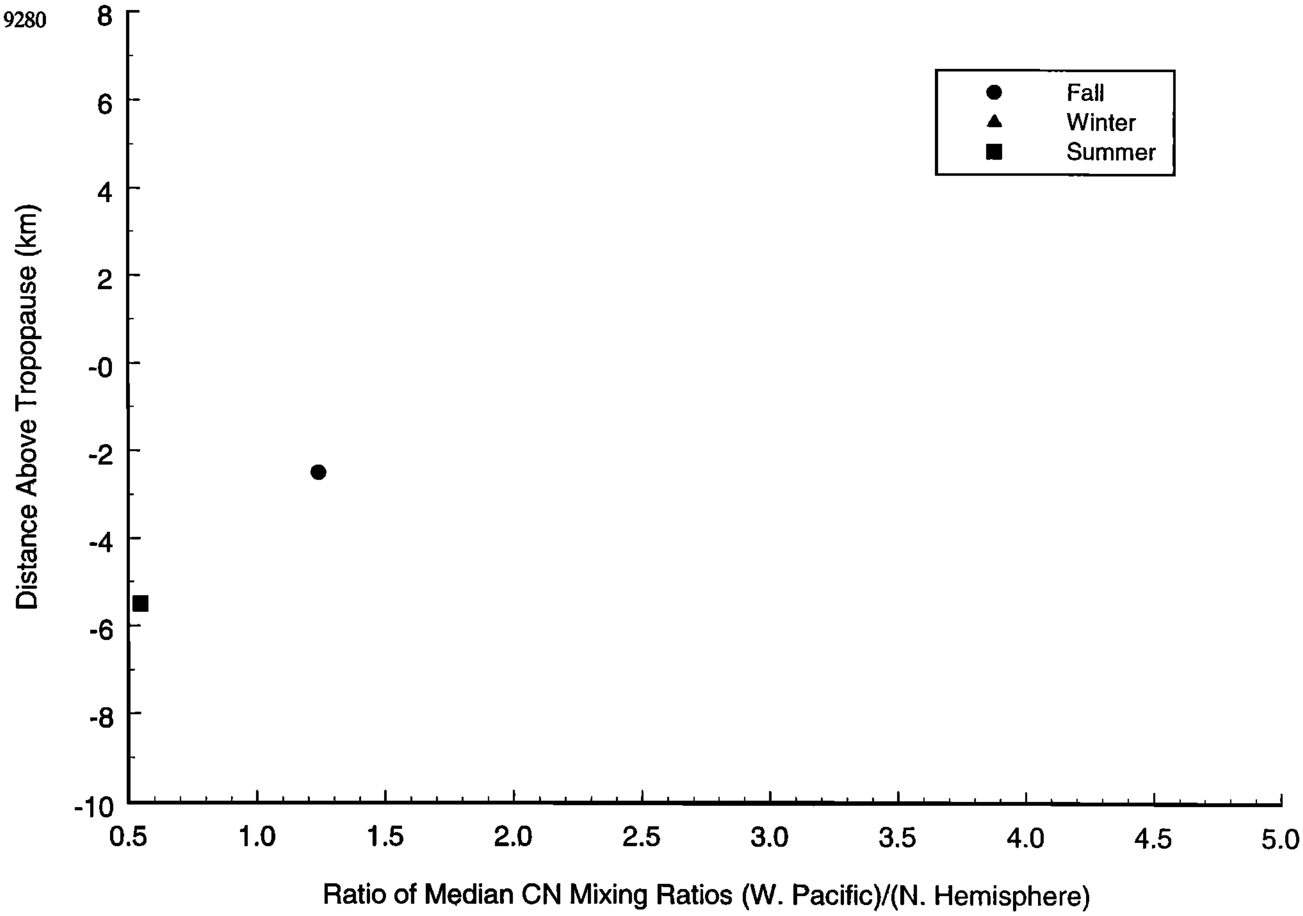


Figure 7. (continued)



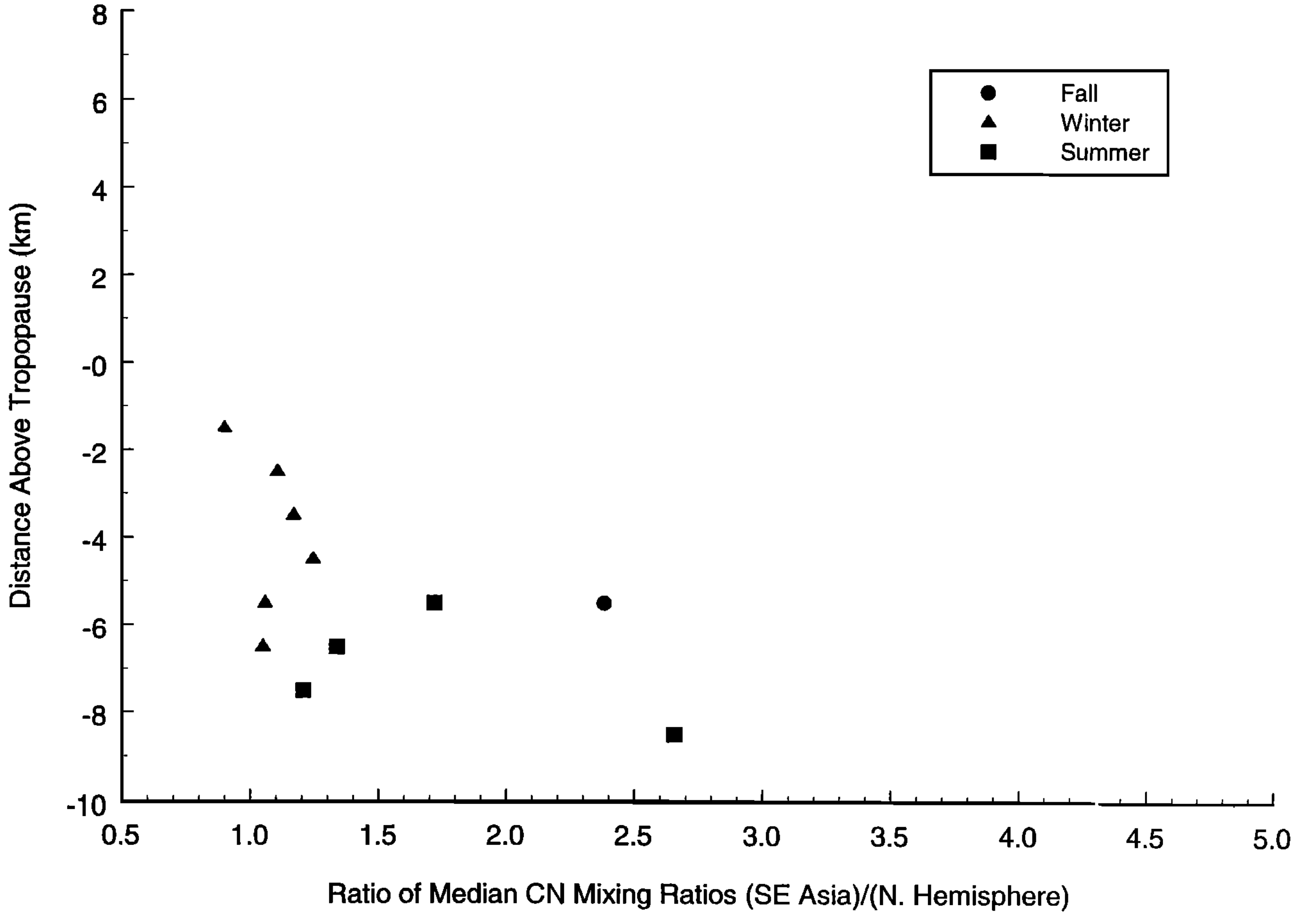


Figure 7. (continued)

more volatile particles preferentially emitted by aircraft engines would yield vertical profiles with different characteristics.

The relatively constant CN number mixing ratio in clear air within the middle and upper troposphere at all levels suggests relatively efficient mixing of CN emitted at the surface, the dominant global source of CN, vertically throughout the troposphere. It is noteworthy that the regions with generally the highest CN mixing ratios in the GASP upper tropospheric observations are southeast Asia, North Atlantic, and Europe. Each of these regions is downwind, in the sense of the prevailing winds in the middle and upper troposphere, of major industrial regions with relatively higher CN emissions at the surface. Southeast Asia and Europe are themselves regions with high surface emissions. Aircraft and surface emissions can lead to higher concentrations at upper levels in downwind regions as well as over the regions where the emissions occurred. While the surface emissions are being carried downwind, they are mixing vertically through the troposphere. Upward and downwind transport is one factor that may explain how relatively higher emissions, both at the surface and over the United States, can lead to higher observed CN mixing ratios in the upper troposphere over the North Atlantic (a region with negligible surface CN emissions) compared to over the United States itself.

These observations can be compared to the estimates, presented by *Kellogg* [1978] of the global geographic distribution of lower tropospheric aerosol in the early to mid 1970s, and to estimates by *Cooke et al.* [1999] of emissions of fine carbon particles by combustion processes. Kellogg's estimates were based on mean wind patterns and the assumption that lower tropospheric aerosol concentration was proportional to the gross national products of countries in a region. *Cooke et al.* [1999] used more quantitative fossil fuel consumption data. The surface patterns of *Kellogg* [1978] and *Cooke et al.* [1999] are quite consistent with those found in the GASP CN data if one allows for slightly more downstream advection in the case of the upper tropospheric/lower stratospheric aerosol than in the case of the boundary layer aerosol.

These observations also are consistent with the more limited and more recent observations of upper tropospheric aerosol reported by *Dibb et al.* [1998]. They reported higher sulfate concentrations over the central United States compared to the eastern Pacific at altitudes below 10 km, but little gradient in sulfate mixing ratios at higher altitudes. At altitudes near and above the tropopause, the relationship between source regions and ambient aerosol mixing ratios can be offset great distances in space due to the horizontal movement of surface emissions during the time they are mixing from source regions at the surface vertically through the troposphere.

7. Conclusions

Worldwide measurements of upper tropospheric/lower stratospheric CN mixing ratios were obtained along commercial aircraft routes from October 1977 through June 1979, during GASP. Comparison of these observations to other CN observations in that era suggests that the GASP expansion-type CN counters responded to fewer CN than the thermal diffusion chamber-based counters flown as part of the University of Wyoming balloonborne instrument packages used to survey tropospheric and stratospheric aerosols and gases. The portions of the GASP observations in clear air show relatively little variation in median mixing ratios in the vertical within the middle and upper troposphere. Mixing ratios decrease with height in the stratosphere, and even in the upper troposphere in some regions and seasons. Median mixing ratios varied between 10^8 and 10^9 CN kg^{-1} in the troposphere, and are near and below 10^8 CN kg^{-1} in the lower stratosphere. The highest mixing ratios in the upper troposphere were observed in the southeast Asia, North Atlantic, and Europe regions, as defined in the *Baughcum et al.* [1996a] aircraft fuel consumption and emission inventories.

The lack of a strong vertical gradient in CN mixing ratios in clear air within the troposphere suggests that both surface and aircraft emissions generally are well mixed in the vertical in clear regions. With the assumption that most anthropogenic CN originate in processes involving fossil fuel combustion, our observations are consistent with the idea that aircraft contributions to global anthropogenic tropospheric CN were in proportion to the ratio of aircraft fuel consumption to total fossil fuel combustion, or of the order of 2%, during the late 1970s. Regional enhancements of up to a factor of 3.4 compared to median Northern Hemisphere CN concentrations in the upper troposphere and lower stratosphere were observed, but variations in regional median CN concentrations did not correlate consistently with regional fuel usage ratios. The regions with highest median mixing ratios generally were also the regions with highest surface and/or aircraft sources, or were just downwind of such high emission regions.

Acknowledgments. It is a pleasure to acknowledge guidance from J. D. Holdeman, G. D. Nastrom, and R. Pratt while getting this study organized, and consultations with A. Hogan on CN counters. W.-Y. Chang helped with the database design. T. Deshler provided the CN data from the University of Wyoming. Many members of the Atmospheric Effects of Aviation Community provided useful feedback to our presentations and questions. This effort was supported by grant NAG5-2711 from NASA, and with funds from the State of South Dakota.

References

Anderson, B.E., W.R. Cofer, D.R Bagwell, J.W. Barrick, C H Hudgins, and K E. Brunke, Airborne observations of aircraft aerosol emissions, 1, Total nonvolatile particle emission indices, *Geophys. Res. Lett*, *25*, 1689-1692, 1998a

Anderson, B E., W R Cofer, J D Barrick, D.R Bagwell, and C.H. Hudgins, Airborne observations of aircraft aerosol emissions, 2, Factors controlling volatile particle production, *Geophys Res Lett*, *25*, 1693-1696, 1998b.

Baughcum, S.L., T.G. Tritz, S C Henderson, and D C. Pickett, Scheduled civil aircraft emission inventories for 1992: Database development and analysis, *NASA Contr. Rep. 4700*, 72 pp., NASA, Washington, D.C., 1996a.

Baughcum, S.L , S.C. Henderson, and T.G. Tritz, Scheduled civil aircraft emission inventories for 1976 and 1984: Database development and analysis, *NASA Contr Rep 4722*, 51 pp., NASA, Washington, D.C., 1996b.

Bodhaine, B.A., Aerosol measurements during the Mauna Loa Photochemistry Experiment 2, *J. Geophys. Res.*, *101*, 14,757-14,765, 1996.

Brasseur, G.P., R A. Cox, D. Hauglustaine, I Isaksen, J. Lelieveld, D.H. Lister, R. Sausen, U. Schumann, A. Wahner, and P. Wiesen, European scientific assessment of the atmospheric effects of aircraft emissions, *Atmos. Environ.*, *32*, 2329-2418, 1998

Briehl, D., In situ measurement of particulate number density and size distribution from an aircraft, *NASA Tech Memo.*, *X-71577*, 1974.

Brown, L.R., M. Renner, and C. Flavin, *Vital Signs 1997, The Environmental Trends That Are Shaping Our Future*, 165 pp., W.W. Norton, New York, 1997.

Brunner, D., J. Staehelin, and D. Jeker, Measurements of nitrogen oxides and ozone from a Swissair B-747 jumbo: Project NOXAR, paper presented at the 1997 Conference on the Atmospheric Effects of Aviation, Virginia Beach, Va., March 10-14, NASA, 1997.

Cadle, R.D., G. Langer, J.B. Haberl, A. Hogan, J.M. Rosen, W.A. Sedlacek, and J. Wegrzyn, A comparison of the Langer, Rosen, Nolan-Pollak and SANDS condensation nucleus counters, *J. Appl. Meteorol.*, *14*, 1566-1571, 1975.

Clarke, A.D., Atmospheric nuclei in the Pacific midtroposphere: Their nature, concentration, and evolution, *J. Geophys. Res.*, *98*, 20,633-20,647, 1993.

Cooke, W F., C. Liousse, H. Cachier, and J. Ferchter, Construction of a 1°x1° fossil fuel emission data set for carbonaceous aerosol and implementation and radiative impact in the ECHAM4 model, *J. Geophys. Res.*, *104*, 22,137-22,162, 1999.

Dibb, J.E., R. W Talbot, and M.B Loomis, Tropospheric sulfate distribution during SUCCESS: Contributions from jet exhaust and surface sources, *Geophys. Res. Lett*, *25*, 1375-1378, 1998.

Fahey, D.W., et al., Emission measurements of the Concorde supersonic aircraft in the lower stratosphere, *Science*, *270*, 70-74, 1995.

Ferry, G.V., R.F. Pueschel, A.W. Strawa, Y. Kondo, S.D. Howard, S. Verma, M.J. Mahoney, T.P. Bui, J.R. Hannan, and H.E. Fuelberg, Effects of aircraft on aerosol abundance in the upper troposphere, *Geophys. Res. Lett.*, *26*, 2399-2402, 1999.

Folkins, I., M. Loewenstein, J. Podolske, S.J. Oltmans, and M. Profitt, A barrier to vertical mixing at 14 km in the tropics: Evidence from ozonesondes and aircraft. *J. Geophys. Res.*, *104*, 22,095-22,102, 1999.

Friedl, R.R., Atmospheric effects of subsonic aircraft: Interim assessment report of the Advanced Subsonic Technology Program, *Rep. NASA RP-1400*, 168 pp., NASA Cent. for Aerosp. Inf., Linthicum Heights, Md., 1997.

Gauntner, D.J , T. Nyland, M. Tiefermann, and T.J. Dudzinski, Measurements of carbon monoxide, condensation nuclei, and ozone on a B 747SP aircraft flight around the world, *Geophys. Res. Lett.*, *6*, 167-170, 1979.

Hagen, D., P. Whitefield, J. Paladino, M. Trueblood, and H. Lilenfeld, Particulate sizing and emission indices for a jet engine exhaust sampled at cruise, *Geophys. Res. Lett.*, *25*, 1681-1684, 1998.

Hofmann, D J., Aircraft sulfur emissions, *Nature*, *349*, 659, 1991.

Hofmann, D J., Twenty years of balloon-borne tropospheric aerosol measurements at Laramie, Wyoming, *J. Geophys. Res.*, *98*, 12,753-12,766, 1993.

Hoffman, D J . R S Stone. M E Wood, T. Deshler, and J.M Harris, An analysis of 25 years of balloon-borne aerosol data in search of a signature of the subsonic commercial aircraft fleet, *Geophys Res Lett.*, *25*, 2433-2436, 1998

Hogan, A W., Some characteristics of Aitken nucleus concentrations, *J Rech. Atmos*, *2*, 87-93, 1966

Hogan, A.W , A transect of the southern circumpolar jet stream, *J. Appl. Meteorol*, *20*, 1239-1241, 1981.

Hogan, A W , Aerosol exchange in the remote troposphere, *Tellus*, *38B*, 197-213, 1986

Hogan, A W., and V.A Mohnen, On the global distribution of aerosols, *Science*, *205*, 1373-1375, 1979

Hogan, A W., W Winters, and S.C Barnard, Photo electric nucleus counters for aerosol climatology, *J. Aerosol Sci.*, *12*, 477-490, 1981.

Holdeman, J.D , F M Humenik, and E.A. Lezberg, NASA Global Atmospheric Sampling Program (GASP) data report for tape VL0004, *NASA Tech. Memo x-73574*, 29 pp., 1976. (Available as *N7713563* from Natl. Tech. Inf. Serv., Springfield, Va.)

Huebert, B.J., G Lee, and W.L. Warren, Airborne aerosol inlet passing efficiency measurement, *J. Geophys. Res*, *95*, 16,369-16,381, 1990.

Jasperson, W H., G D. Nastrom, and J.D. Holdeman, Climatology of ozone from 19000 to 59000 feet based on combined GASP and ozonesonde data. *NASA Tech. Pap.*, *TP-2303*, 1984.

Jasperson, W H., G.D. Nastrom, R.E. Davis, and J D. Holdeman, Variability of cloudiness at airline cruise altitudes from GASP measurements, *J Clim. Appl. Meteorol.*, *24*, 74-82, 1985.

Kellogg, W W., Global influences of mankind on climate, in *Climatic Change*, edited by J. Gribbin, pp. 205-227, Cambridge Univ. Press, New York, 1978.

Lezberg, E A., F M. Humenik, and D. Otteson, Sulfate and nitrate mixing ratios in the vicinity of the tropopause, *Atmos. Environ.*, *13*, 1299-1304, 1979.

Marenco, A , et al., Measurement of ozone and water vapor by Airbus in-service aircraft· The MOZAIC airborne program, An overview, *J Geophys. Res*, *103*, 25,631-25,642, 1998.

Miake-Lye, R.C , B.E. Anderson, W.R. Cofer, H.A. Wallio, G.D. Nowicki, J.O. Ballenthin, D. E. Hunton, W.B. Knighton, T.M. Miller, J.V. Seeley, and A.A. Viggiano, SOx oxidation and volatile aerosol in aircraft exhaust plumes depend on fuel-sulfur content, *Geophys. Res. Lett*, *25*, 1677-1680, 1998.

Nastrom, G.D., Ozone in the upper troposphere from GASP data, *J. Geophys. Res.*, *84*, 3683-3688, 1979.

Nastrom, G.D., and D.C. Fritts, Sources of mesoscale variability of gravity waves, 1 Topographic excitation, *J. Atmos. Sci.*, *49*, 101-110, 1992.

Nastrom, G.D., and K.S. Gage, A climatology of atmospheric wave number spectra of wind and temperature observed by commercial aircraft, *J. Atmos Sci*, *42*, 950-960, 1985.

Nastrom, G.D., J.D. Holdeman, and R.E. Davis, Cloud encounter and particle number density variabilities from GASP data, *J. Aircraft*, *19*, 272-277, 1982.

Newell, R.E., and D.J. Gauntner, Experimental evidence of interhemispheric transport from airborne carbon monoxide measurements, *J Appl. Meteorol.*, *18*, 696-699, 1979.

Nyland, T.W., Condensation-nuclei (Aitken particle) measurement system used in the NASA Global Atmospheric Sampling Program, *NASA Tech. Pap.*, *1415*, 27 pp., 1979.

Paladino, J., P. Whitefield, D. Hagen, A.R. Hopkins, and M. Trueblood, Particle concentration characterization for jet engine emissions under cruise conditions, *Geophys. Res. Lett.*, *25*, 1697-1700, 1998.

Perkins, P J., and U.R.C. Gustafsson, An automated sampling system operating on 747 airliners, *Rep. NASA TMX-71790*, 8 pp., NASA, Washington, D.C., 1975.

Pratt, R., and P. Falconer, Circumpolar measurements of ozone, particles, and carbon monoxide from a commercial airliner, *J. Geophys. Res.*, *84*, 7876-7882, 1979.

Pruppacher, H.R., and J.D. Klett, *Microphysics of Clouds and Precipitation*, 287 pp., Kluwer Acad., Norwell, Mass., 1997.

Pueschel, R , S. Verma, G.V. Ferry, S.D. Howard, S. Vay, S. Kinne, J. Goodman, and A. Strawa, Sulfuric acid and soot particle formation in aircraft exhaust, *Geophys. Res. Lett.*, *25*, 1685-1688, 1998.

Ridley, B A., E L. Atlas, J G. Walega, G. L Kok, T A. Staffelbach, J P. Greenberg, F. E Grahek, P G. Hess, and D. D Montzka, Aircraft measurements made during the spring maximum of ozone over Hawaii: Peroxides, CO, O_3, NO_y, condensation nuclei, selected hydrocarbons, halocarbons, and alkyl nitrates between 0 5 and 9 km altitude, *J Geophys. Res.*, *102*, 18,935-18,961, 1997.

Rosen, J.M., and D.J. Hofmann, Balloon-borne measurements of condensation nuclei, *J Appl Meteorol*, *16*, 56-62, 1977.

Schlager, H , P. Konopka, P. Schulte, U. Schumann, H. Ziereis, F Arnold, M. Klemm, D.E. Hagen, P D. Whitefield, and J. Ovarlez, In situ observations of air traffic emission signatures in the North Atlantic flight corridor, *J. Geophys Res.*, *102*, 10,739-10,750, 1997.

Schroder, F., and J. Strom, Aircraft measurements of sub micrometer aerosol particles (>7 nm) in the midlatitude free troposphere and tropopause region, *Atmos. Res*, *44*, 333-356, 1997.

Schumann, U., J. Strom, R. Busen, R. Baumann, K. Gierens, M. Krautstrunk, F.P. Schroder, and J Stingl, In situ observations of particles in jet aircraft exhausts and contrails for different sulfur-containing fuels, *J. Geophys Res.*, *101*, 6853-6869, 1996.

Skala, G.F., A new instrument for the continuous measurement of condensation nuclei, *Anal Chem.*, *35*, 702-706, 1963.

Spicer, C.W., M.W. Holdren, D.L. Smith, D.P. Hughes, and M.D. Smith, Chemical composition of exhaust from aircraft turbine engines, *J. Eng. Gas Turbines Power*, *114*, 111-117, 1992.

Spicer, C.W., M.W. Holdren, R.M Riggin, and T F. Lyon, Chemical composition and photochemical reactivity of exhaust from aircraft turbine engines, *Ann Geophys.*, *12*, 944-955, 1994.

Strom, J., and S. Ohlsson, In situ measurements of enhanced crystal number densities in cirrus clouds caused by aircraft exhaust, *J. Geophys. Res.*, *103*, 11,355-11,361, 1998.

Toon, O.B., and R.C. Miake-Lye, Subsonic aircraft: Contrail and cloud effects special study (SUCCESS), *Geophys. Res. Lett.*, *25*, 1109-1112, 1998.

Weber, R.J., A D. Clarke, M. Litchy, J. Li, G. Kok, R.D. Schillawski, and P.H. McMurry, Spurious aerosol measurements when sampling from aircraft in the vicinity of clouds. *J. Geophys. Res.*, *103*, 28,337-28,346, 1998.

Whitefield, P.D., M.B. Trueblood, and D. E. Hagen, Size and hydration characteristics of laboratory simulated jet engine combustion aerosols, *Part. Sci. Technol.*, *11*, 25-36, 1993.

Wozniak, A.G., Exploratory analysis of Global Atmospheric Sampling Program data, M. Sc. dissertation, 173 pp., S.D. Sch. of Mines and Technol., Rapid City, 1997.

Wu, M.-F., Analysis of GASP carbon monoxide data. *NASA Conf Rep.*, *CR-165365*, 1981.

A.G. Detwiler and L.R. Johnson, Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, Rapid City, SD 57701. (andy@ias.sdsmt.edu)

A.G. Schauer, Science and Engineering Services, Inc., NASA Wallops Flight Facility, Bldg. N-159 Rm. E206, Wallops Island, VA 23336.

(Received March 26, 1999; revised December 2, 1999;
accepted December 9, 1999.)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9283–9290, APRIL 16, 2000

Supercooling of type 1 polar stratospheric clouds: The freezing of submicron nitric acid aerosols having HNO_3 mol fractions less than 0.5

A. K. Bertram, D. B. Dickens, and J. J. Sloan

Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada

Abstract. We report the freezing temperatures for nitric acid/water aerosol particles having radii about 0.2 μm and concentrations in the range $0.24 < x\mathrm{HNO_3} < 0.46$ ($x\mathrm{HNO_3}$ = nitric acid mol fraction). The aerosols supercool by up to 85 K. The droplets having $x\mathrm{HNO_3} = 0.33$ have the highest freezing point in this concentration range; they supercool by 60 K and produce nitric acid dihydrate (NAD) upon freezing. We describe a simplified principal components analysis that improves the detection of the freezing point and allows the identification of the solid that precipitates upon freezing. This procedure shows that for the temperature range and experimental conditions explored, only NAD precipitates in the concentration range $0.33 < x\mathrm{HNO3} < 0.42$ and both NAD and nitric acid trihydrate (NAT) precipitate for the range $0.24 < x\mathrm{HNO_3} < 0.33$. The temperature dependences of our measured freezing points correlate strongly with NAD saturation ratios, but have no relationship with NAT supersaturation at any concentration. Thus NAD nucleates preferentially from supercooled aerosols in this composition range, probably because of kinetic constraints on the nucleation of NAT. We conclude that NAD nucleation is possible during rapid cooling events in the polar stratosphere, if temperatures lower than 185 K are reached. During the freezing of stratospheric aerosol droplets in this concentration range, it seems likely that NAD nucleates first, thereby providing a surface on which NAT may crystallize.

1. Introduction

It is generally believed that polar stratospheric clouds (PSCs) are composed of ternary mixtures of H_2SO_4, HNO_3, and H_2O and that the ratios of these components may vary considerably depending on the ambient stratospheric conditions. At temperatures well above the ice frost point, the particles are believed to be composed of sulfuric acid and water with only a small amount of nitric acid. When the temperature approaches the ice frost point, however, a considerable amount of nitric acid and water vapor condenses on the sulfate aerosol, and the fraction of H_2SO_4 in the ternary particle becomes very small [*Molina et al.*, 1993]. Field measurements [Beyerle et al., 1997] and equilibrium models [*Carslaw et al.*, 1994; *Tabazadeh et al.*, 1994] suggest that more than 90% of the gas phase HNO_3 condenses at temperatures close to the frost point, forming particles in which the H_2SO_4 component can be less than 5 wt %.

This condensation process is most pronounced at the extremely low temperatures found in mountain lee waves, where very large increases in PSC volume have been observed [*Deshler et al.*, 1994]. These orographically induced temperature decreases also can be so rapid that they cause diffusional disequilibrium, further enhancing the HNO_3 concentrations in smaller particles and decreasing the H_2SO_4 fraction until it approaches 0.01 wt % [*Meilinger et al.*, 1995; *Tsias et al.*, 1997]. Under these extreme conditions it seems likely that PSCs consist of near-binary HNO_3/H_2O particles.

Both solid and liquid PSC particles have been detected in the stratosphere by lidar [*Browell et al.*, 1990; *Toon et al.*, 1990] and balloon-borne sondes [*Larsen et al.*, 1997]. Despite a large body of observational and laboratory data, however, the phases of PSC particles under a given set of stratospheric conditions cannot yet be predicted a priori [*Peter*, 1997]. Kinetic barriers to nucleation and crystal growth may cause both the compositions and phases to deviate substantially from thermodynamic equilibrium, so in many cases of stratospheric interest, the particles may be in a metastable form. The phases of PSC particles govern their effectiveness in chlorine activation [*Ravishankara and Hanson*, 1996] so it is important to determine the conditions under which phase changes occur.

We have undertaken a series of laboratory measurements to determine the freezing behavior of nitric acid aerosols having a range of compositions that might be found in the polar stratospheric regions. In two previous publications we reported the freezing of 2:1 H_2O:HNO_3 molar ratio aerosol droplets to nitric acid dihydrate (NAD) [*Bertram and Sloan*, 1998a] and 3:1 aerosol droplets to nitric acid trihydrate (NAT) [*Bertram and Sloan*, 1998b]. In that previous work we reported both the freezing temperature for the aerosol droplets and the temperature-dependent nucleation rate constants for the hydrates. The purpose of the present paper is to extend our measurements to cover nitric acid mol fractions between 0.25 and 0.5. For this composition range we will report the temperature for the onset of freezing, and additionally we will identify directly the crystalline hydrate that precipitates first. In many cases this is different from the corresponding results predicted by the bulk phase diagram.

2. Measurements

The measurement is designed to determine the temperature at which crystalline material nucleates in micron-sized droplets of aqueous nitric acid having compositions relevant to those of type

Copyright 2000 by the American Geophysical Union.

Paper number 1999JD901174.
0148-0227/00/1999JD901174$09.00

1 PSCs. The techniques and apparatus have been described previously [*Bertram and Sloan*, 1998a, b]. The experiments were carried out in a temperature-programmable flow tube consisting of three copper sections, thermally isolated from each other by thin-walled stainless steel bellows. Each section can be independently cooled to about 120 K. Temperature differences between adjacent sections can be as much as 100 K. The axial temperature profile is determined by 10 copper/constantan thermocouples fixed to the walls of the copper sections. Three are located in each of the first two sections, and four are in the third (observation) section. In separate experiments the detailed temperature structure of the tube was mapped using thermocouples mounted on movable stainless steel probes. These latter measurements showed the temperature of the gas relaxes rapidly to that of the wall on moving from one section to another. For the flow rates used in our experiments, about 4 standard liters per minute (slpm), complete equilibration occurs in approximately the first inch of the second and third copper sections (vide infra). The residence times in the first two sections were 11 s, and in the last section it was 25 s.

Vapor phase H_2O and HNO_3, mixed with nitrogen carrier gas, are admitted to the first section of the tube, which is held at 188 K. These immediately condense to micron-sized liquid droplets; no solid nucleates at this temperature. The resulting aerosol flows through the second and third sections, which are programmed to lower temperatures. The IR spectra of the particles are recorded through windows located at the end of the third section, in order to detect the first appearance of crystalline material. The criteria for the identification of the crystalline features in the IR spectra will be discussed in detail in the Results section.

The temperature programming required in the second and third sections depends on the material being observed and the information sought. For cases where the crystal growth rate is rapid, or if it is only necessary to detect the temperature at which crystals first appear in the droplet, then the temperature of the second section is held at 188 K, and the temperature of the third section is lowered slowly, resulting in a two-stage temperature profile along the tube. If it is necessary to ensure that the entire droplet freezes, and the crystal growth rate is slow at low temperatures, then a three-stage temperature profile is used. In this case, the temperature of the second section is lowered below 188 K to induce nucleation, and the temperature of the third section is held at 188 K to ensure rapid crystallization of the droplets that nucleate in the second section. Both temperature programs were used in the work to be reported here.

The freezing temperatures are very sensitive to the acid concentration in the aerosol droplets. For the present experiments the concentrations were determined by revaporizing the particles after they left the flow tube and measuring the spectrum of the vapor as it passed through a glass cell at the exit of the flow tube using a second Fourier transform infrared (FTIR) spectrometer. This procedure assumes that the vapor pressures of both HNO_3 and H_2O are sufficiently low at the temperature of the flow tube that all the vapor entering the glass cell comes from revaporized aerosol, a condition that is met for all the experiments reported here. Absolute concentrations were determined by recording spectra of HNO_3 and H_2O vapor in equilibrium with aqueous nitric acid solutions of known concentrations. The total pressures of nitrogen carrier gas in the calibration measurements were the same as those in the experiment. The HNO_3 and H_2O partial pressures in the carrier gas were calculated using the equilibrium model of [*Taleb et al.*, 1996]. As a check on the accuracy of this procedure, we used it to determine the concentrations of 2:1 and 3:1 aerosols that had been established using the spectroscopic methods outlined in our previous papers on NAD and NAT. Both methods gave concentrations that agreed within +/- 0.01 in terms of the nitric acid mol fraction. Freezing experiments were carried out covering the range of HNO_3 mol fractions ($x\mathrm{HNO_3}$) from $0.24 \leq x\mathrm{HNO_3} \leq 0.46$.

3. Results

The freezing behavior of the nitric acid aerosol particles varies strongly with concentration. During the limited observation time available with our apparatus, some concentrations do not crystallize at all, while others crystallize completely. For those cases in which crystallization occurs, we wish to determine the temperature at which freezing begins and the identity of the hydrate that first precipitates. If possible, we also wish to determine the changes in the composition of the particles as they solidify. We have developed a spectral decomposition procedure based on a simplified form of principal component analysis to get this information. We describe this procedure first, then present the results.

According to the nitric acid phase diagram, only four phases can be present during the freezing of a particle in this concentration range: NAT, NAD, nitric acid monohydrate (NAM), and liquid. [*Barton et al.*, 1993] have shown that NAM does not crystallize in submicron aerosols on a time scale of minutes, and our observations, presented below, confirm this. Consequently, for temperatures at and just below that at which the first solid appears, only NAD, NAT, and liquid can be present. We make the assumption that the spectrum of the aerosol droplets during freezing can be described as a linear combination of the spectra of these three components (the basis spectra). Within certain limits that we will discuss shortly, this assumption appears to be valid. We calculated the basis spectra of the hydrates using optical constants for NAT and NAD obtained from those provided by R. E. Miller et al. [see *Niedziela et al.*, 1998]. As the basis spectrum of the liquid, we chose the spectrum of the liquid aerosol droplets, recorded just before the first appearance of any solid.

Defining the system in this way permits us to represent the problem as a set of simultaneous linear equations with unknown coefficients. Such sets of equations can be solved using simple matrix algebra. The unknown coefficients, which give the amounts of each component in the droplet, are the matrix product of the inverse of the array of basis spectra times the measured spectrum.

The approximations involved in this treatment all relate to the assumption that the basis spectra are invariant during the freezing of the particle. The validity of this assumption depends on two things. For the spectra of the solids the changes in light scattering due to particle growth must be negligible. For the liquid spectrum the concentration changes due to the deposition of the hydrates must be negligible. We now explore the conditions under which these assumptions are valid.

The effect of the change in size of the solid particles can be inferred from the spectra plotted in Figure 1. This shows Mie theory calculations, using the optical constants of [*Niedziela et al.*, 1998], of the spectra of NAD particles having lognormal size distributions with a standard deviation of 1.6 and geometric radii of 0.2 μm (top) and 0.02 μm (bottom). These radii bracket those

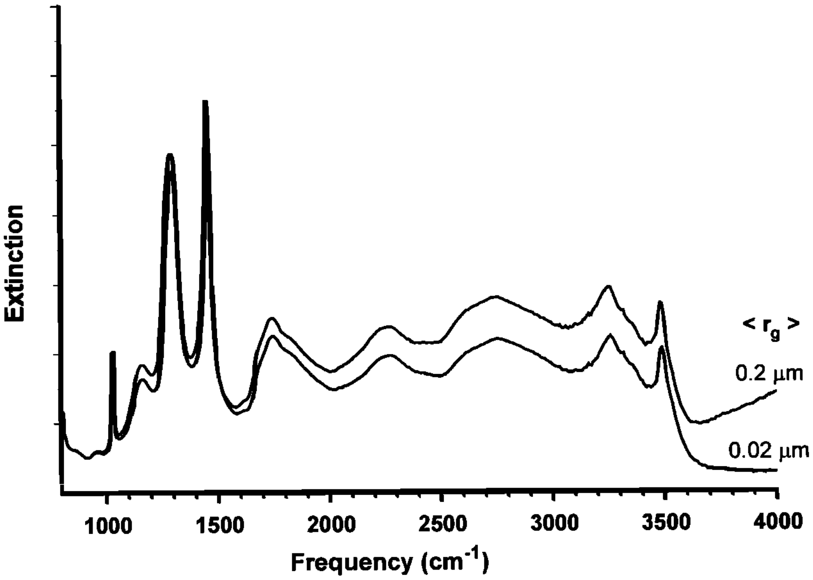


Figure 1. Mie theory calculations of the spectra of small NAD particles showing that the effects of scattering are not significant in the indicated size range.

in our experiment, while the value of σ is typical for our conditions. The amplitudes have been normalized at the intense peak near 1450 cm^{-1}. In our analysis procedure we used only that part of the spectrum between 1000 and 2000 cm^{-1}, for which the change in scattering due to particle size causes a shift of 4% in the relative spectral amplitudes. Note that this small change comes from a change of 3 orders of magnitude in the particle volume and hence the spectrum amplitude. In an experiment for which the signal to noise of the larger crystals is just over 100, the smaller ones are not detectable, so the maximum error introduced by changes in light scattering during the deposition of the solid in our experiment should be less than 4%.

The effects of neglecting the changes in liquid concentration are illustrated in Figure 2. This shows spectra of liquids having $x\mathrm{HNO_3}$ = 0.46 (lowest, heavy line); 0.33 (middle, dashed line), and 0.24 (top, thin line). The spectra were recorded just above the temperature at which the solid first appeared in the particle and are normalized at the peak near 1300 cm^{-1}. The concentrations bracket those used in the present work. For the concentration range 0.46 < $x\mathrm{HNO_3}$ < 0.33 the spectra are nearly identical; the intensity differences over the entire range shown are less than 10%. When the mol ratio of H_2O to HNO_3 increases above 2:1 ($x\mathrm{HNO_3}$ < 0.33), the intensity of the OH band near 3300 cm^{-1} increases significantly, while that of the nitrate bands between 1000 cm^{-1} and 1500 cm^{-1} does not. If this OH band be excluded by considering only the range from 1000 to 2000 cm^{-1}, however, it is found that the maximum intensity difference is 20% (at 1450 cm^{-1}) and the average difference in spectral

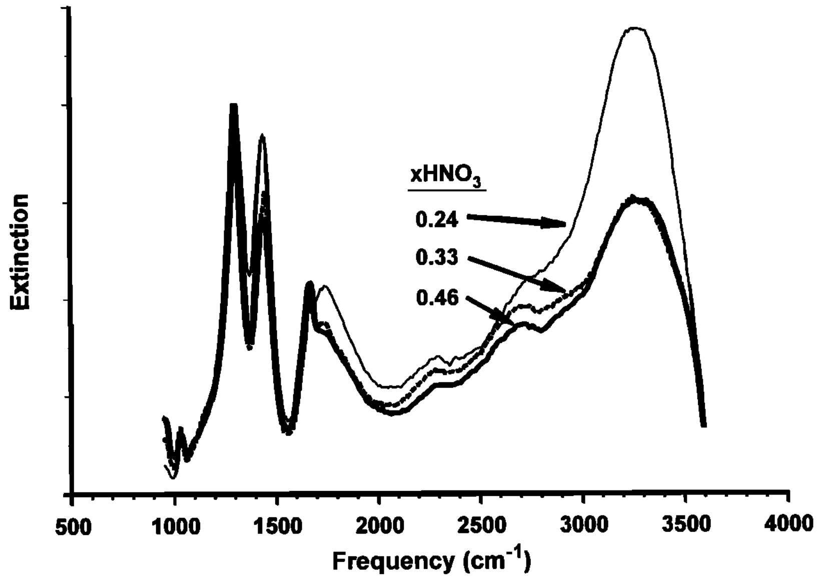


Figure 2. Measured spectra of liquid aerosol droplets having radii of approximately 0.2 μm and indicated HNO_3 concentrations. The spectra between 1000 cm^{-1} and 2000 cm^{-1} are not strongly concentration-dependent.

intensities over this range is less than 5%. Limiting the spectral bandwidth to the range between 1000 and 2000 cm^{-1} therefore provides a reasonable compromise between the conflicting requirements of adequate spectral bandwidth to detect the changes, and invariability of the liquid spectrum with composition. This range was used for the spectrum decomposition calculations.

On the basis of the above argument we conclude that the decomposition procedure based on the spectral range from 1000 to 2000 cm^{-1} is valid for the composition range $0.46 < x\mathrm{HNO_3} < 0.24$. The largest error is introduced by the changes in the liquid spectrum during freezing of the more dilute aerosols and within the errors quoted above, this procedure can determine the composition of the nonstoichiometric aerosol particles during the entire freezing process from liquid to solid. Note that as the amount of liquid becomes smaller and hence its composition differs more from that of the original liquid, its contribution to the infrared signal will diminish as well, often becoming less than the noise before the particle has completely frozen.

The above argument notwithstanding, we will not attempt to report the composition during the entire freezing process, but only to detect the first appearance of solid in the droplet and to determination its identity (NAD or NAT). When used only for these purposes, the change in concentration of the liquid is negligible, so the variation shown in Figure 2 is not relevant. This procedure is used to determine the freezing temperatures reported in the following sections. For this purpose, the onset of freezing is defined to be the temperature at which the fraction of liquid in the droplet decreases below 0.95 for the indicated concentration. Note that this differs from the procedure used in our previous papers on the freezing of NAT and NAD, where we reported the freezing point as the temperature of the first spectrum after the point where the difference between certain bands in successive spectra changes from zero (see *Bertram and Sloan* [1998a, b] for details). We believe that the present method in principle should be a more sensitive way to detect the freezing point since it uses a larger part of the spectrum.

3.1. Concentration Range: $0.46 \geq x\mathrm{HNO_3} \geq 0.42$

Spectra of aerosols in this composition range were recorded over the temperature range from 186 K to 156 K using both two-stage and three-stage temperature profiles. In both cases the spectra over the entire temperature range were those of either liquid or amorphous nitric acid; no crystallization could be detected. (Note that the liquid and amorphous spectra are so similar that we cannot distinguish between them.) Freezing of the bulk liquid in this concentration range produces NAM. Comparison of the observed aerosol spectra with that of NAM, however, showed no evidence that it is formed in the aerosol. Virtually identical spectra to ours have been observed previously in this temperature range for $\mathrm{HNO_3}$ aerosols [*Barton et al.*, 1993] and thin films [*Koehler et al.*, 1992]. We conclude therefore that the homogeneous nucleation for this composition range is too slow to be observed in our experiment, even at a temperatures of 156 K, which represents supercooling of more than 80 K. We identify nucleation as the rate-limiting process rather than crystal growth because the same result was observed using a three-stage temperature profile. Two stage temperature profiles were used for the measurements in the more dilute concentration ranges, which will be reported in the next two sections. Several of these were also checked using three stage profiles. Within the error bars (see Discussion and Figure 5 below) the two methods gave the same temperature for the appearance of the first solid (onset of freezing).

3.2. Concentration Range: $0.42 \geq x\mathrm{HNO_3} \geq 0.33$

The temperatures for the onset of freezing (as defined above) vary from about 174 K to 183 K, depending on concentration. Figure 3 shows a series of spectra for aerosols having $x\mathrm{HNO_3} = 0.38$ in the temperature range between 169 K and 186 K. The spectra are approximately constant for temperatures above and below this range, but they begin to change rapidly as the temperature decreases below about 178 K, due to the precipitation of a crystalline solid. We used the spectral decomposition

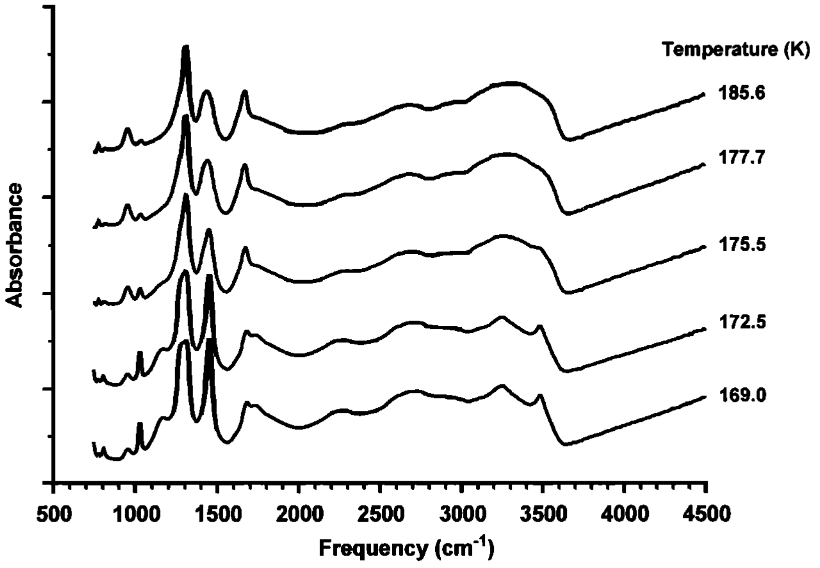


Figure 3. Temperature dependence of aerosol spectra for $x\mathrm{HNO_3} = 0.38$.

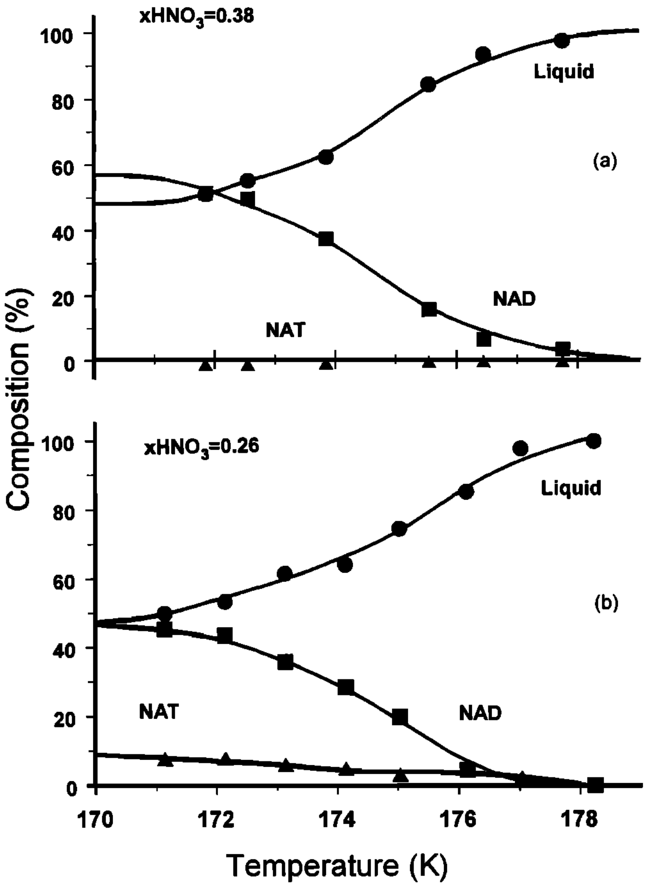


Figure 4. Composition of aerosol droplets during freezing for initial $x\mathrm{HNO_3}$ = (a) 0.38 and (b) 0.26. Two-stage temperature profiles were used in both of these measurements.

procedure with these spectra to identify the temperature where the first solid appears, and also to identify this solid. The liquid basis spectrum used for the analysis at $x\mathrm{HNO_3}$ = 0.38 was the top one in Figure 3 (recorded at 186 K). The NAD and NAT basis spectra were computed from the optical constants of R. E. Miller et al. (vide supra) using a temperature of 160 K and a lognormal size distribution with r_g = 0.2 µm and σ = 1.6. These parameters gave the best fits to the NAD and NAT spectra recorded in experiments on the freezing of droplets with 2:1 and 3:1 stoichiometry under the same experimental conditions. (We comment in the Discussion on the uncertainties in r_g and σ stemming from the lack of uniqueness in these fits.)

The coefficients derived from the spectral decomposition for $x\mathrm{HNO_3}$ = 0.38 are shown in Figure 4a as percentage compositions. The temperature at which the liquid component drops below 95%, the freezing point, is 177 ± 1.6 K in this case. The error is from the uncertainty in the temperature measurement. The coefficients in Figure 4a show that crystalline NAD is the first solid formed on freezing for this concentration. The coefficient for the NAT contribution (triangles) remains at zero within the estimated error of ± 5% for the entire temperature range. The NAT coefficient becomes slightly negative at lower temperatures, indicating that the basis spectra change slightly at these low temperatures. All of the experiments on aerosol droplets with mol fractions between 0.42 and 0.33 gave negligible NAT coefficients throughout the entire temperature range. We conclude that in this concentration range, NAD is the only crystalline hydrate that precipitates on freezing. At temperatures substantially below the freezing point, growth of NAD crystals slows due to increasing liquid viscosity.

3.3. Concentration Range: 0.33 > $x\mathrm{HNO_3} \geq$ 0.25

NAD was the first and also the predominant hydrate that precipitated for most of this concentration range, but NAT also appeared as a minor component, which became predominant in the immediate vicinity of $x\mathrm{HNO_3}$ = 0.25. The coefficients for the case $x\mathrm{HNO_3}$ = 0.26 are shown in Figure 4b. The contribution from NAT is small near the freezing point, but rises to about 10% at low temperatures. The freezing temperature at this concentration is low, and approximately equal to that of the higher concentration shown in Figure 4a. The freezing temperature reaches a maximum between these compositions, near $x\mathrm{HNO_3}$ = 0.33 and drops thereafter as the composition approaches 0.25.

Since NAT is the more stable hydrate at these temperatures, it might be expected that NAT would constitute a large fraction of the solid phase as the liquid composition approaches 3:1 $\mathrm{H_2O{:}HNO_3}$. The fact that the metastable NAD is formed in preference to NAT (using two-stage temperature profiles) indicates differences in the kinetics of nucleation. The former must have a lower barrier to nucleation than the latter.

4. Discussion

The freezing point results are summarized in Figure 5. The top part of the figure shows the equilibrium phase diagram of the $\mathrm{HNO_3}$ – $\mathrm{H_2O}$ system [*Ji et al.*, 1996]. The experimentally determined aerosol freezing temperatures are shown in the lower panel. Note the break in the temperature axis. The circular points are the aerosol freezing temperatures determined in the series of experiments described above, using the spectral decomposition procedure. The solid curve is a nonlinear least squares fit to these points. The estimated errors in the individual composition/temperature determinations are indicated by the error bars on the point for $x\mathrm{HNO_3}$ = 0.32; these provide a guide to the uncertainty in the measurements carried out in this series of experiments.

The two square points are the freezing temperatures determined for the $x\mathrm{HNO_3}$ = 0.25 and 0.33 aerosols reported in our two previous papers on NAT and NAD [*Bertram and Sloan*, 1998a, b]. These are lower than the present measurements by about 3 K. Although part of this difference might be attributable to the temperature uncertainty of our measurement, ± 1.6 K, we attribute it mainly to differences in the particle sizes stemming from changes we made to the methods used for the two series of experiments. Diffusion of $\mathrm{H_2O}$ to the flow tube walls is a source of difficulty for experiments of this kind because it leads to instability in the aerosol concentration. To prevent this, we increased the flow rates for the present experiments by about a factor of 2, from 2.5 slpm to about 4 slpm, and increased the water vapour flow to compensate for this change. This appears to have both decreased the average particle size and increased the width of the size distribution, with the result that there were some larger particles in the present experiments. We have not

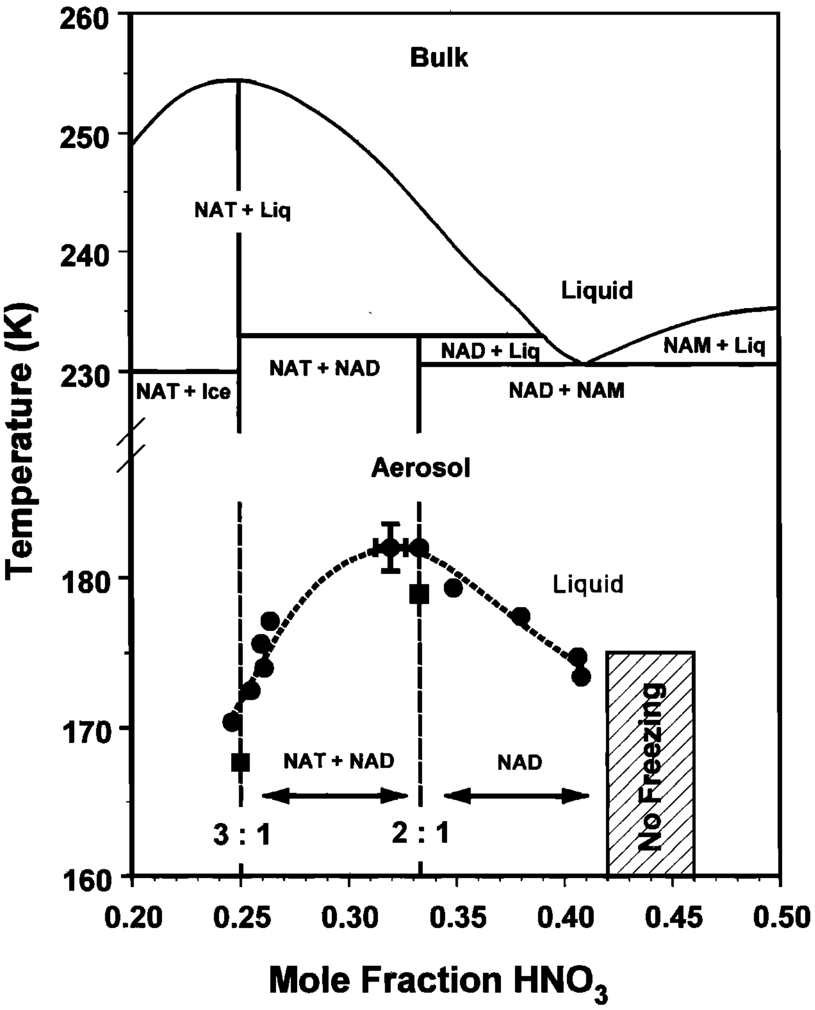


Figure 5. Kinetic phase diagram showing freezing temperatures of nitric acid aerosols compared to bulk behavior. The larger barrier to NAT nucleation is evident from the lower freezing temperatures near $x\text{HNO}_3 = 0.25$.

quantified the change in the size distribution accurately due to the nonuniqueness of the size distributions obtained from the Mie calculations, to which we referred in our previous work [*Bertram and Sloan*, 1998b]. In classical nucleation theory, however, the nucleation rate constant is proportional to the exponential of the cube of the radius, so the nucleation rate is a sensitive function of particle size. In an experiment such as ours, this has the result that the nucleation and freezing of the larger particles is detected by the spectral decomposition procedure at a higher temperature, leading to an upward shift in the freezing curve. The size sensitivity of the nucleation rate effects the temperatures at which different aerosol size distributions freeze, both in laboratory and in the stratosphere, so measured freezing temperatures are expected to vary among different experiments. Since the present experiments were all carried out using the same procedure, the relative temperatures at which the different concentrations freeze are accurate to within the errors quoted above, and are the significant result that we wish to discuss below.

Clearly, there is a large difference between the bulk solid-liquid equilibrium curves and the freezing curve of the aerosol. The latter curves are between 60 and 85 K lower than the bulk equilibrium curves. More importantly, the shapes of these curves are quite different. In the bulk equilibrium case the highest freezing temperature occurs for $x\text{HNO}_3 = 0.25$, which freezes at approximately 11 K above the solution for which $x\text{HNO}_3 = 0.33$. For the case of the aerosol, $x\text{HNO}_3 = 0.25$ freezes about 10 K lower than $x\text{HNO}_3 = 0.33$. This result is independent of the method used to detect the freezing and of the number of stages (two or three) in the temperature profile. These experiments allow us to conclude that the crystal growth rate is not responsible for the observed difference; it is due solely to the activation barrier to nucleation.

The coefficients obtained in the spectral decomposition analysis give the identity of the solid components that first appear on freezing. In Figure 4 we showed that for $x\text{HNO}_3 = 0.38$, NAD is the first solid that precipitates, whereas for $x\text{HNO}_3 = 0.26$, both NAT and NAD appear. These results occur consistently in the respective composition ranges. Only NAD precipitates from droplets having $x\text{HNO}_3 > 0.33$, whereas NAD and NAT both appear in the spectra from droplets having $0.25 < x\text{HNO}_3 < 0.33$. These identifications are indicated in the appropriate concentration ranges in the lower half of Figure 5.

At temperatures below about 170 K the kinetic barrier to crystal growth becomes significant. We demonstrate the effect of this barrier in Figure 6, which shows the result of two separate freezing measurements on aerosols having $x\text{HNO}_3 = 0.25$. In the

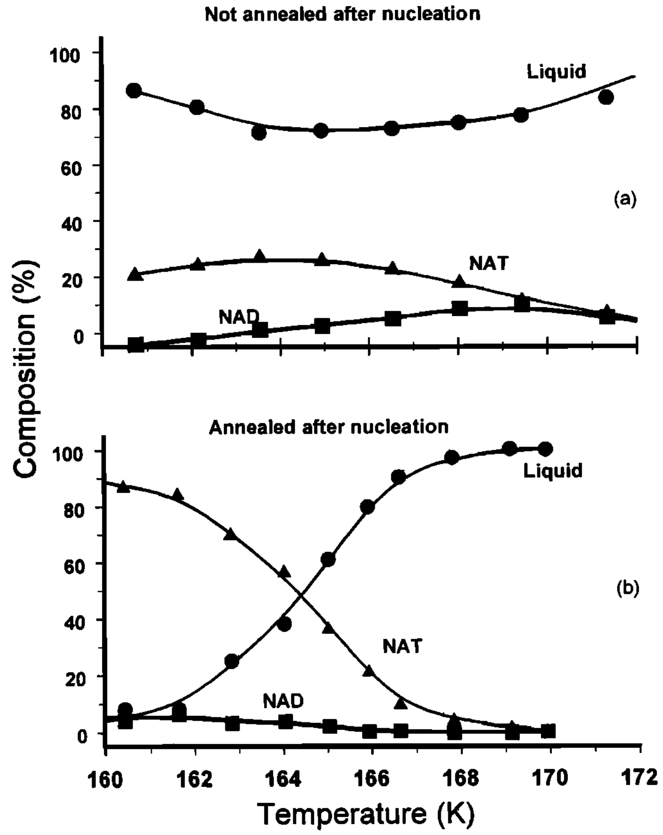


Figure 6. Composition of aerosol droplets having xHNO$_3$ = 0.25 during freezing. (a) Droplet temperature is decreased monotonically. (b) Droplet temperature is reduced, then increased to 188 K to enhance crystal growth.

top panel, only two temperature steps were used. The droplets were formed at 188 K in the first section, the second section was also held at 188 K, and the temperature was reduced to the value shown on the x axis in the third section. Small amounts of both NAT and NAD are formed at very low temperatures, but no substantial crystallization of either occurs. In the lower panel a similar experiment was carried out with the exception that only the temperature in the second section was reduced, while the third section was held at 188 K in order to facilitate the growth of crystals nucleated in the second section. This procedure gives NAT as the predominant component, as would be expected from relative thermodynamic stabilities. In this case as well, the major difference in the onset temperature between this and the previous work is likely due to differences in the particle size distribution.

The fact that both NAT and NAD form in aerosols with xHNO$_3$ ranging from 0.33 to 0.25 raises the question of which hydrate nucleates first. We examined this via calculations of the NAD and NAT saturation ratios in the liquid solutions. The saturation ratios of NAT and NAD, S_{NAT} and S_{NAD}, are defined as

$$S_{NAT}=(p\mathrm{HNO_3})(p\mathrm{H_2O})^3/K_{NAT}, \qquad (1)$$

$$S_{NAD}=(p\mathrm{HNO_3})(p\mathrm{H_2O})^2/K_{NAD}, \qquad (2)$$

where pHNO$_3$ and pH$_2$O are the equilibrium vapor pressures of nitric acid and water over the liquid solutions and K_{NAT} and K_{NAD} are equilibrium constants. These saturation ratios were calculated using the thermodynamic model and the temperature-dependent activity product of NAT reported by *Carslaw et al.*, [1995], and the temperature-dependent activity product of NAD reported by *Massucci et al.*, [1999]. The results of these calculations are shown in Figure 7. Clearly, there is a strong correlation between the observed freezing temperatures and the NAD saturation ratios. On freezing, the aerosols in our experiments were all supersaturated with respect to NAD by about the same amount, by factors of 20 to 24. The NAT saturation ratios, however, do not correlate with the observed freezing temperatures and are at least twice as large for much of the concentration range. This is strong evidence that homogeneous nucleation of supercooled nitric acid aerosols produces NAD, rather than NAT for xHNO$_3$ in the range from 0.25 to 0.42.

On the basis of the above evidence we conclude that NAD is the first solid that nucleates from the aerosol having xHNO$_3$ = 0.25, despite the fact that at the observed freezing temperature, S_{NAT} is approximately 60 whereas S_{NAD} is only 22. Thus the barrier to NAT nucleation must be substantially larger than that to NAD nucleation. A similar conclusion has been drawn previously from other experimental evidence [*Barton et al.*, 1993; *Bertram and Sloan*, 1998b; *Prenni et al.*, 1998; *Worsnop et al.*, 1993]. If it is indeed the case that NAD always nucleates first in these aerosols, then the nucleation rate constant reported by *Bertram and Sloan*, [1998b] is that for NAD nucleation from xHNO$_3$ = 0.25 and subsequent crystal growth of NAT on this nucleus, as shown in Figure 6b. It is important both for the physical chemistry of the nitric acid system and for the atmospheric implications that this point be resolved.

5. Atmospheric Implications

As mentioned in the introduction, liquid aerosol droplets in the polar stratosphere can reach near-binary HNO$_3$-H$_2$O concentrations having sulfuric acid concentrations less than 5 wt

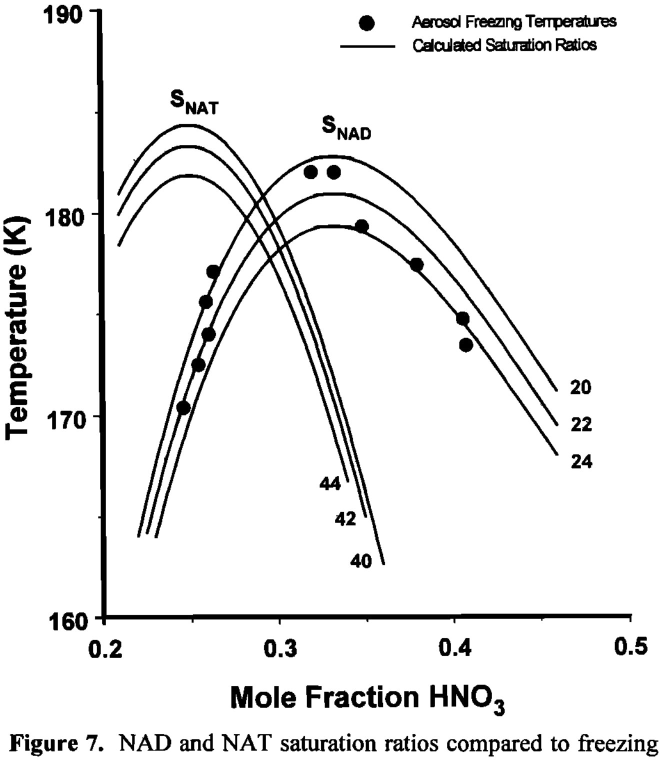


Figure 7. NAD and NAT saturation ratios compared to freezing temperatures observed in the present experiments for the indicated droplet concentrations.

%. *Koop et al.*, [1995] have demonstrated that these particles remain liquid until temperatures below the ice frost point are reached. More recent theoretical work [*Meilinger et al.*, 1995; *Tsias et al.*, 1997] has suggested that during rapid temperature fluctuations the sulfuric acid concentration in small liquid droplets may approach 0.01 wt % and that the nitric acid mol fractions may approach values of 0.24 to 0.28. As an example, we consider the freezing of 1% of a sample of 0.2 μm diameter particles during a rapid temperature fluctuation lasting 5 min. For this to occur, the homogeneous nucleation rate must be about $10^9 cm^{-3}s^{-1}$. On the basis of the residence times and particle sizes for the measurements shown in Figure 5, we can estimate that the homogeneous nucleation rate constants in our experiments are approximately $1\text{-}3\times10^{10} cm^{-3}s^{-1}$. These nucleation rate constants are highly temperature-dependent. If we use the temperature dependence reported in our previous publications on NAD and NAT to extrapolate them to higher temperatures, we find that a nucleation rate of $10^9 cm^{-3}s^{-1}$ occurs between 170 and 185 K; that is, 1% of a sample of 0.2 μm diameter particles with concentrations in the range $0.24 < x\mathrm{HNO_3} < 0.28$ would nucleate in 5 min at these temperatures. These temperatures are lower than those used by Meilinger et al. and Tsias et al. in their modeling studies, and they suggest that very low temperatures would be required to nucleate NAD.

Acknowledgments. We are pleased to acknowledge financial support for this work from the Natural Sciences and Engineering Research Council and Atmospheric Environment Service Canada. A.K.B. acknowledges the receipt of an NSERC postgraduate scholarship.

References

Barton, N., B. Rowland, and J.P. Devlin, Infrared spectra of large acid hydrate clusters: Formation conditions of submicron particles of nitric acid dihydrate and nitric acid trihydrate, *J. Phys. Chem.*, *97*, 5848-5851, 1993.

Bertram, A.K., and J.J. Sloan, Temperature-dependent nucleation rate constants and freezing behavior of submicron nitric acid dihydrate aerosol particles under stratospheric conditions, *J. Geophys. Res.*, *103*, 3553-3561, 1998a.

Bertram, A.K., and J.J.Sloan, The nucleation rate constants and freezing mechanism of nitric acid trihydrate aerosol under stratospheric conditions, *J. Geophys. Res.*, *103*, 13,261-13,265, 1998b.

Beyerle, G., et al., Temperature dependence of ternary solution particle volumes as observed by lidar in the Arctic stratosphere during winter 1992/1993, *J. Geophys. Res.*, *102*, 3603-3609, 1997.

Browell, E.V., et al., Airborne lidar observations in the wintertime Arctic stratosphere: Polar stratospheric clouds, *Geophys. Res. Lett.*, *17*, 385-388, 1990.

Carslaw, K.S., et al., Stratospheric aerosol growth and HNO_3 gas phase depletion from coupled HNO_3 and water uptake by liquid particles., *Geophys. Res. Lett.*, *21*, 2479-2482, 1994.

Carslaw, K.S., S.L.Clegg, and P.Brimblecombe, A thermodynamic model of the system $HCl\text{-}HNO_3\text{-}H_2SO_4\text{-}H_2O$, including solubilities of HBr, *J. Phys. Chem.*, *99*, 11,557-11,574, 1995.

Deshler, T., et al., Lifetime of leewave-induced ice particles in the Arctic stratosphere; 1, Balloonborne observations, *Geophys. Res. Lett.*, *21*, 1327-1330, 1994.

Ji, K., et al., The use of X-ray diffraction in support of calorimetry for the characterization of some nitric acid hydrates possibly involved in the polar stratospheric clouds formation, *Geophys. Res. Lett.*, *23*, 981-984, 1996.

Koehler, B.G., A.M. Middlebrook, and M.A.Tolbert, Characterization of model polar stratospheric cloud films using Fourier transform infrared spectroscopy and temperature programmed desorption, *J. Geophys. Res.*, *97*, 8065-8074, 1992.

Koop, T., et al., Do stratospheric aerosol droplets freeze above the ice frost point?, *Geophys. Res. Lett.*, *22*, 917-920, 1995.

Larsen, N., et al., Temperature histories in liquid and solid polar stratospheric cloud formation, *J. Geophys. Res.*, *102*, 23,505-23,517, 1997.

Massucci, M., S.L. Clegg, and P. Brimblecombe, Equilibrium partial pressures, thermodynamic properties of aqueous and solid phases, and Cl_2 production from aqueous Hcl and HNO_3 and their mixtures, *J. Phys. Chem. A*, *103*, 4209-4226, 1999.

Meilinger, S.K., et al., Size-dependent stratospheric droplet composition in lee wave temperature fluctuations and their potential role in PSC freezing, *Geophys. Res. Lett.*, *22*, 3031-3034, 1995.

Molina, M.J., et al., Physical chemistry of the H2SO4/HNO3/H2O system: implications for polar stratospheric clouds, *Science*, *261*, 1418-1423, 1993.

Niedziela, R.F., R.E. Miller, and D.R. Worsnop, Temperature- and frequency-dependent optical constants for nitric acid dihydrate from aerosol spectroscopy, *J. Phys. Chem. A*, *102*, 6477-6484, 1998.

Peter, T., Microphysics and heterogeneous chemistry of polar stratospheric clouds, *Annu. Rev. Phys. Chem.*, *48*, 785-822, 1997.

Prenni, A.J., et al., Composition-dependent freezing nucleation rates for HNO3/H2O aerosols resembling gravity wave perturbed stratospheric particles, *J. Geophys. Res.*, *103*, 28,439-28,450, 1998.

Ravishankara, A.R., and D.R. Hanson, Differences in the reactivity of type-I polar stratospheric clouds depending on their phases, *J. Geophys. Res.*, *101*, 3885-3890, 1996.

Tabazadeh, A., et al., A study of type I polar stratospheric cloud formation, *Geophys. Res. Lett.*, *21*, 1619-1622, 1994.

Taleb, D.-E., J.-L. Ponche, and P. Mirabel, Vapor pressures in the ternary system water-nitric acid-sulfuric acid at low temperature: A reexamination, *J. Geophys. Res.*, *101*, 25,967-25,977, 1996.

Toon, O.B., et al., An analysis of lidar observations of polar stratospheric clouds, *Geophys. Res. Lett.*, *17*, 393-396, 1990.

Tsias, A., et al., Freezing of polar stratospheric clouds in orographically induced strong warming events, *Geophys. Res. Lett.*, *24*, 2303-2306, 1997.

Worsnop, D.R., et al., Vapor pressures of solid hydrates of nitric acid: Implications for polar stratospheric clouds, *Science*, *259*, 71-74, 1993.

A. K. Bertram, D. B. Dickens, and J. J. Sloan, Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. (sloan@uwaterloo.ca)

(Received September 1, 1999; revised November 30, 1999; accepted December 7, 1999.)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9291-9302, APRIL 16, 2000

Statistical characterization of air ion mobility spectra at Tahkuse Observatory: Classification of air ions

U. Hõrrak, J. Salm, and H. Tammet

Institute of Environmental Physics, University of Tartu, Tartu, Estonia

Abstract. A database of 8615 hourly averaged air ion mobility spectra in the range of 0.00041–3.2 cm^2 V^{-1} s^{-1} was measured at Tahkuse Observatory, Estonia, during 14 months in 1993–1994. The average mobility spectrum over the whole period shows distinct peaks of small and large ions. Intermediate ions with mobilities of 0.034–0.5 cm^2 V^{-1} s^{-1} are of low concentration of about 50 cm^{-3} in the average spectrum. They experience occasional bursts of up to about 900 cm^{-3} during 6–10 hours at daytime. The number of burst events recorded during 14 months was 101, with maximum frequency in spring and minimum frequency in winter. Physically, large and intermediate ions can be called aerosol ions, and small ions can be called cluster ions. The principal component analysis was applied to detect the structure of an air ion mobility spectrum. As a result, the mobility spectrum in the range of 0.00041–3.2 cm^2 V^{-1} s^{-1} (diameters of 0.36–79 nm) was divided into five classes: small cluster, big cluster, intermediate, light large, and heavy large ions. The boundaries between the classes are 1.3 cm^2 V^{-1} s^{-1} (diameter of 0.85 nm), 0.5 cm^2 V^{-1} s^{-1} (1.6 nm), 0.034 cm^2 V^{-1} s^{-1} (7.4 nm), and 0.0042 cm^2 V^{-1} s^{-1} (22 nm). The five principal components that are closely correlated with the respective ion classes explain 92% of total variance. The classification of aerosol ions is in accord with the three-modal structure of the size spectrum of submicron aerosol particles.

1. Introduction

Measurements of the mobility spectra of natural air ions could be most generally characterized by the mobility range and resolution and by the frequency and duration of recordings. In various papers, these characteristics have varied to a large extent, depending on particular goals and technical resources of the researchers. The spectrometer designed by *Misaki* [1961a] has a high resolution of eight logarithmically divided fractions per decade of mobility. At first the spectra of small ions in the range of 0.2–3 cm^{2}V^{-1}s^{-1} were measured at two different sites in Japan during a few days [*Misaki*, 1961b]. Later the spectra of large ions in the range of 0.00018–0.01 cm^{2}V^{-1}s^{-1} (11 days) and in the range of 0.000042–0.0024 cm^{2}V^{-1}s^{-1} (3 days) were measured in the New Mexico semidesert in 1963 [*Misaki*, 1964]. Thereafter measurements in the wide range of 0.0001–3.2 cm^{2}V^{-1}s^{-1} were carried out at three sites in Japan; the whole duration of the measurements was about 1 month [*Misaki et al.*, 1972]. *Kojima* [1984] measured air ion mobility spectra in the range of 0.0085–0.24 cm^{2}V^{-1}s^{-1}. Five series (7–10 days) of measurements were carried out at the campus of the Science University of Tokyo in Noda during three seasons from summer 1983 to spring 1984. *Dhanorkar and Kamra* [1991, 1993a] designed and built a mobility spectrometer with three measuring condensers that covers a range of 0.00023–3.4 cm^{2}V^{-1}s^{-1}. They recorded 28 spectra (6 spectra a day) at Pune, India, in 1991 [*Dhanorkar and Kamra*, 1993a]. The concentrations of small, intermediate, and large ions were recorded at the same place during nearly 1 year in 1990–1991 [*Dhanorkar and Kamra*, 1993b].

Owing to the complexity and large-scale variability of atmospheric processes, episodic measurements are not sufficient to characterize the regularities of the mobility spectra of natural air ions. Long-term measurements of air ions in a wide range of mobility are necessary to draw statistically founded conclusions about the shape and variations of the mobility spectra for periods of different duration.

The classification of air ions represents one essential problem that can be studied by long-term measurements of air ion spectra. The classification has been established gradually [*Israël*, 1970; *Flagan*, 1998], but it has not been satisfactorily formulated until now. The concepts of small and large ions have a clear physical background [*Tammet*, 1995]. Problems arise when trying to specify the concept of intermediate ions and settle the mobility boundaries. The boundaries defined in atmospheric electricity textbooks are rather speculative conventions. One way of development is the statistical analysis of the air ion spectra measured in a wide mobility range, in order to search for air ion groups with different statistical properties. A natural classification should explain the coherent behavior of air ions inside class intervals and the relative independence of the ions of different classes. Measurements used in the verification of the classification are required to record air ion mobility fractions that are narrow in comparison with mobility classes. The analysis of the statistical behavior of fraction concentrations requires thousands of mobility spectra recorded during at least one full year. The first measurements that allow statistical classification of air ions were carried out at Tahkuse Observatory.

Copyright 2000 by the American Geophysical Union.

Paper number 1999JD901197
0148-0227/00/1999JD901197$09.00

The measurement of detailed mobility spectra in natural atmospheric air at Tahkuse Observatory has been running since 1985. The mobility spectrum of small air ions together with a narrow fraction of light intermediate ions was measured from June 1985 to June 1986 [*Hõrrak et al.*, 1988]. Instrumentation for measurements in a wide mobility range was set into operation in July 1988. A brief summary of measurements for the period until 1989 was reported by *Hõrrak et al.* [1994]. A description of the behavior of intermediate ions at Tahkuse Observatory for the period from September 1993 to October 1994 was presented by *Hõrrak et al.* [1998b]. A statistical synopsis of the air ion spectra for the entire mobility range at the same place and for the same period is given in the present paper, with emphasis on the study of air ion classification. The large and intermediate ions are charged aerosol particles. Thus the problem is related to the size classification of atmospheric aerosol particles.

2. Measurements

2.1. Location

Tahkuse Observatory with coordinates 58°31'N, 24°56'E is located in a sparsely populated rural region. It is 27 km northeast of the city of Pärnu and 100 km south of Tallinn, the capital of Estonia. Pärnu, with 52,000 inhabitants, is located on the coast of the Gulf of Riga, at the east coast of the Baltic Sea. The terrain surrounding the observatory consists of flat open country with some tree groups (about 100 trees in a radius of 100 m), small woods, grassland, and agricultural land. The river of Pärnu is 50 m to the northwest; the nearest neighboring farm is about 200 m west. A road with little automobile traffic passes about 180 m east from the measurement point. The average traffic frequency was about 10 motor vehicles per day, mainly from 0700 to 1900 local standard time (LST), in 1993–1994. The *Soomaa National Park* (Swampland) extends at distances from 6 to 30 km southeast. The weather in this region is quite unsteady owing to the action of cyclones and anticyclones.

2.2. Instrumentation

A complex of air ion spectrometers covering a mobility range of 0.00041–3.2 cm^2 V^{-1} s^{-1} was installed at Tahkuse in 1988 [*Hõrrak et al.*, 1990; *Tammet*, 1990]. The upper mobility limit was chosen to collect the smallest existing air ions. The lower mobility limit is determined by the technical parameters of the equipment. The complex consists of three original multichannel aspiration spectrometers designed according to the principle of the second-order differential mobility analyzer [*Tammet*, 1970]. The spectrometers are by convention called small ion spectrometer (IS$_1$), intermediate ion spectrometer (IS$_2$), and large ion spectrometer (IS$_3$). The illustration of the IS$_1$ and the measuring system is presented in Figure 1. The design of the IS$_2$ and IS$_3$ is analogous. The whole range of mobility is logarithmically divided into 20 intervals (see Table 1): 9 intervals in the subrange of 0.00041–0.29 cm^2 V^{-1} s^{-1} and 11 intervals in the subrange of 0.25–3.2 cm^2 V^{-1} s^{-1}. Thus each mobility spectrum consists of 20 fraction concentrations. The symbols of fractions are N_k and P_k for negative and positive polarity, respectively. The whole mobility range of intermediate ions is covered by means of two spectrometers (IS$_1$ and IS$_2$) of different resolving power. Accordingly, the measured logarithmically distributed fractions of heavy intermediate ions (12–13) are about 3 times wider than those of light intermediate ions (fractions 9–11). The eleventh fraction (0.251–0.320 cm^2 V^{-1} s^{-1}) is overlapped by the twelfth fraction (0.150–0.293 cm^2 V^{-1} s^{-1}).

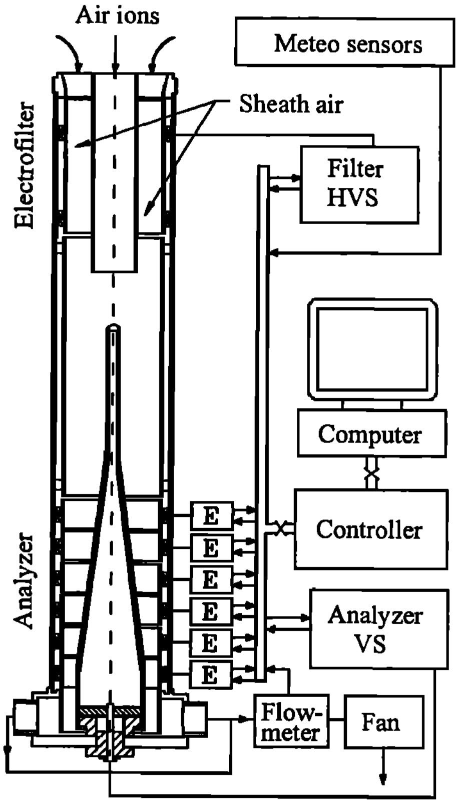


Figure 1. Small air ion spectrometer IS$_1$ and the measuring system. E, electrometer amplifier; HVS, high-voltage supply; and VS, voltage supply. External dimensions of the spectrometer: height, 695 mm; diameter, 122 mm.

The mobility spectra of positive and negative air ions were measured every 5 min. The hourly averages and standard deviations of air ion fraction concentration inside the hourly periods were recorded together with the values of wind direction, wind speed, atmospheric pressure, temperature, relative humidity, and the concentration of NO_2.

The air is sucked into the mobility spectrometers through an opening in the south gable of the building at a height of about 5 m from the ground. To prevent the effect of wind to the airflow, the air inlet (above) and outlet (beneath the inlet) are placed in the same gable with a space of about 1 m. The length of aluminum tube that conducts the air sample to the spectrometers is about 2 m, with a crosssection of 18 × 20 cm^2. There are thin longitudinal metal sheets in the tube for suppression of turbulence. The total air flow rate is about 0.016 m^3 s^{-1}, and air speed is about 0.45 m s^{-1}. The devices, excluding meteorological sensors, are enclosed in a thermally insulated stable-climate chamber, which makes it possible to use the equipment through all the seasons. The chamber and the tube of the air channel are electrically earthed.

2.3. Database

The present paper is based on data collected during the period from September 1, 1993, to October 27, 1994. The period

Table 1. Air Ion Fractions, Estimates of Equivalent Diameter Ranges Assuming Single Charged Particles, and Proposed Classes of Air Ions

Analyzer	Fraction	Mobility $cm^2 V^{-1} s^{-1}$	Diameter nm
		Small Cluster Ions	
IS_1	N_1/P_1	2.51–3.14	0.36–0.45
IS_1	N_2/P_2	2.01–2.51	0.45–0.56
IS_1	N_3/P_3	1.60–2.01	0.56–0.70
IS_1	N_4/P_4	1.28–1.60	0.70–0.85
		Big Cluster Ions	
IS_1	N_5/P_5	1.02–1.28	0.85–1.03
IS_1	N_6/P_6	0.79–1.02	1.03–1.24
IS_1	N_7/P_7	0.63–0.79	1.24–1.42
IS_1	N_8/P_8	0.50–0.63	1.42–1.60
		Intermediate Ions	
IS_1	N_9/P_9	0.40–0.50	1.6–1.8
IS_1	N_{10}/P_{10}	0.32–0.40	1.8–2.0
IS_1	N_{11}/P_{11}	0.25–0.32	2.0–2.3
IS_2	N_{12}/P_{12}	0.150–0.293	2.1–3.2
IS_2	N_{13}/P_{13}	0.074–0.150	3.2–4.8
		Light Large Ions	
IS_2	N_{14}/P_{14}	0.034–0.074	4.8–7.4
IS_2	N_{15}/P_{15}	0.016–0.034	7.4–11.0
IS_3	N_{16}/P_{16}	0.0091–0.0205	9.7–14.8
IS_3	N_{17}/P_{17}	0.0042–0.0091	15–22
		Heavy Large Ions	
IS_3	N_{18}/P_{18}	0.00192–0.00420	22–34
IS_3	N_{19}/P_{19}	0.00087–0.00192	34–52
IS_3	N_{20}/P_{20}	0.00041–0.00087	52–79

under analysis involves 10,224 hours. Owing to occasional pauses in measurements and instrumentation failures, about 16% of the possible measuring time was lost, and 8615 hourly mobility spectra of both signs are available for statistical analysis. The computer program *Statistica for Windows* (*Statsoft Inc.*, 1998) was used for statistical data analysis. A specific *Pascal* program was compiled for the principal component and factor analysis. The recorded air ion mobility fractions and estimates of the equivalent diameter ranges of air ion mobility assuming single charged particles [*Tammet*, 1995, 1998] are presented in Table 1. Five classes of air ions established by means of statistical analysis are also given.

A simplified method for the calculation of fraction concentrations, which does not take into account the shape of apparatus function, was applied in most sections of this paper. The average spectra in Figure 2 were obtained in a stricter way by calculating first the parameters of a piecewise linear spectrum model [*Tammet*, 1980]. The simplified method yields somewhat smoothed mobility spectra. However, the differences, as compared to the stricter method, are small; uncertainties not exceeding a few percent are expected for fraction concentrations [*Tammet et al.*, 1987]. The corrections of the diffusion losses of air ions on the entrance channel parts of spectrometers have been made by relevant equations [*Tammet*, 1970]. The correction factors are $1/(1 - 0.2k^{0.67})$ and $1/(1 - 0.08k^{0.67})$, where k is the mobility of ions, for spectrometer IS_1 and for IS_2+IS_3, respectively.

3. Results and Discussion

3.1. Mobility Spectrum of Air Ions

3.1.1. Average spectra. The average mobility spectra of air ions for the whole period are presented in Figure 2. There are two wide spectral groups with the mobility ranges of 0.5–3.2 and 0.00032–0.034 $cm^2 V^{-1} s^{-1}$, which are traditionally called small ions and large ions, respectively. More detailed average spectra of small ions are presented in Figure 3. The corresponding particle diameters, derived from the electrical mobilities, are presented in the figures assuming single charged particles [*Tammet*, 1995]. The third group lies between large and small ions, with the mobility range of 0.034–0.5 $cm^2 V^{-1} s^{-1}$, and is called intermediate ions. This group appears from time to time as burst events, occasionally occurring around local noon, and its average concentration is about 50 cm^{-3}. Physically, large and intermediate ions may be called aerosol ions, and small ions may be called cluster ions [*Hõrrak et al.*, 1994].

The general shape of the mobility spectra of negative and positive small ions is astonishingly similar to that observed by *Misaki* [1976], whose modes of small ion mobility spectra, 1.56 $cm^2 V^{-1} s^{-1}$ and 1.26 $cm^2 V^{-1} s^{-1}$, are close to those pre-

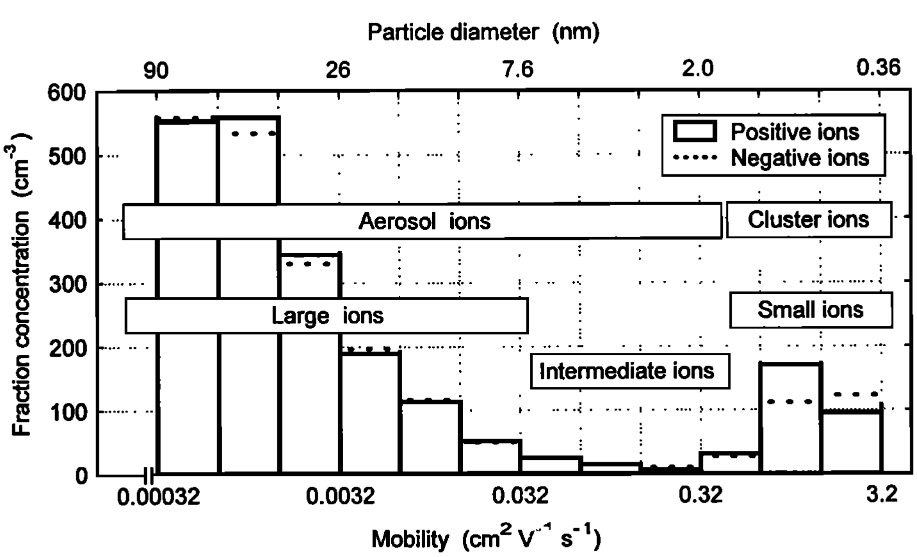


Figure 2. Average mobility spectra of air ions at Tahkuse Observatory, September 1, 1993, to October 27, 1994.

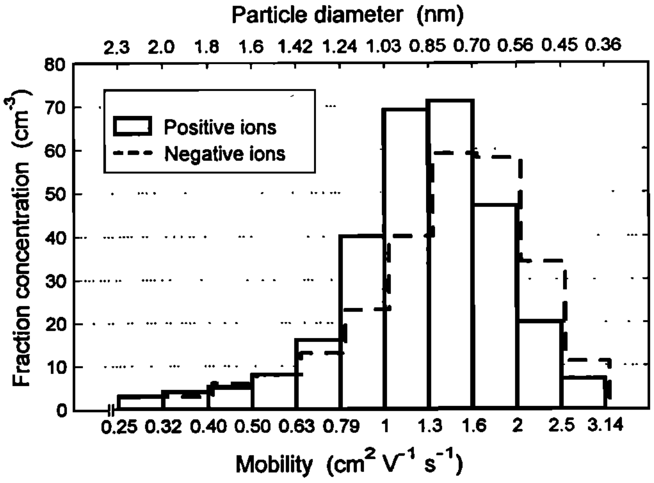


Figure 3. Average spectra of small (cluster) ions at Tahkuse Observatory, September 1, 1993, to October 27, 1994.

sented in Figure 3. Small (cluster) ions are formed in charged state and evolved via ion-molecule reactions in the atmosphere before they obtain their final size [*Mohnen*, 1977; *Luts and Salm*, 1994; *Luts*, 1995; *Nagato and Ogawa*, 1998]. The growth of small ions is thermodynamically hindered at a mobility of 0.5 cm^2 V^{-1} s^{-1} (1.6 nm) in ordinary conditions.

The overall shape of the average spectra in the range of large ions (aerosol ions) is in accord with calculations based on the theory of bipolar charging of aerosol particles by small air ions [*Salm*, 1988; *Hõrrak et al.*, 1998a]. The concentration of large ions diminishes toward higher mobilities owing to the reduction of charging probability and the concentration of aerosol particles. The lower boundary of the spectrum at a mobility of 0.00032 cm^2 V^{-1} s^{-1} is determined by technical limitations of the spectrometer IS$_3$.

The time variations of the air ion mobility spectrum and the aerosol particle size spectrum are well correlated in a size range of 10–80 nm. The correlation coefficient varies from 0.91 to 0.97, depending on the size fraction. These aerosol particles in weakly polluted rural air are believed to be in a quasi-steady charging state [*Hõrrak et al.*, 1998c].

The electrical state of aerosol particles in the intermediate ion range (nanometer particles) is not well known in natural atmosphere. The estimates of charging probability obtained by theoretical considerations and laboratory experiments vary from about 0.5% to 5% for particles from 2 to 10 nm, respectively [*Hoppel and Frick*, 1986; *Reischl et al.*, 1996]. Experimental investigation of competitive ion-induced and binary homogeneous nucleation in gas mixtures shows that the above values may be greatly modified when ions are involved in the nucleation process [*Kim et al.*, 1997, 1998].

3.1.2. Variability of spectra. The relative standard deviation (coefficient of variation) of the hourly averaged values of fraction concentrations is about 50% for small (cluster) air ions and 70% for large air ions. The average fraction concentrations of intermediate ions are relatively low, but their standard deviations are high, up to 130%, owing to the burst events with concentrations up to 900 cm^{-3} [*Hõrrak et al.*, 1998b]. The enhanced concentrations of intermediate ions are recorded from 1000 to 1900 LST, with a duration of 6–10 hours, in fine weather conditions. The relative standard deviations of the fractions of air ion mobility spectra in the daytime (0800–2000 LST) and nighttime (2000–0800 LST) are presented separately in Figure 4. A change of scale is set at a mobility of 0.32 cm^2 V^{-1} s^{-1}, according to technical limitations of the spectrometers, and for better resolution of the spectral regions of aerosol ions and cluster ions.

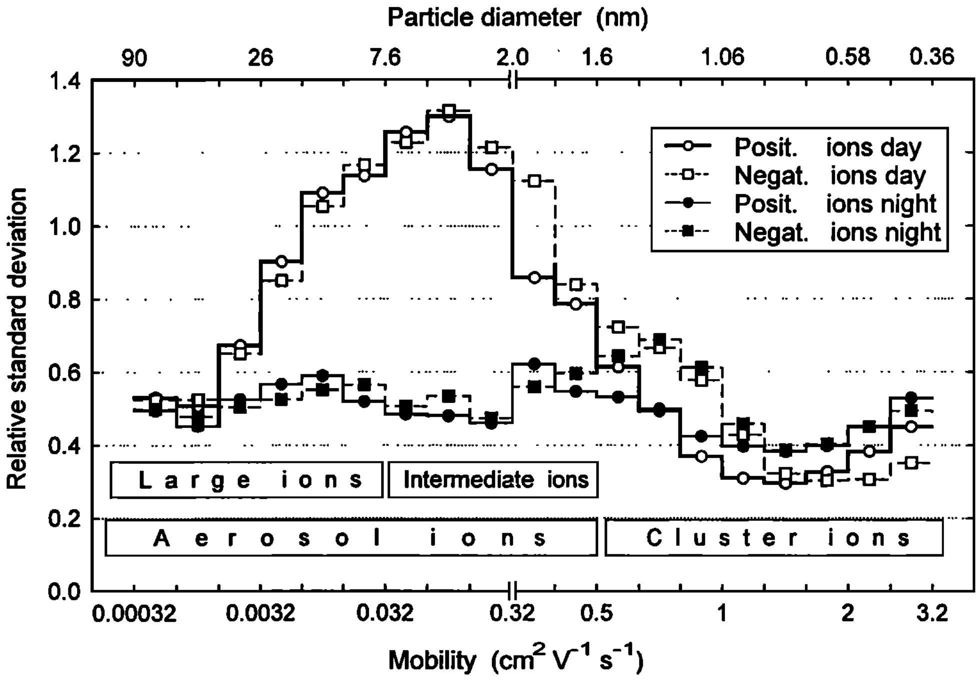


Figure 4. Relative standard deviations of the concentration of air ion spectral fractions at daytime (0700 to 2000 LST) and in the nighttime (2000–0700 LST) at Tahkuse Observatory, September 1, 1993, to October 27, 1994.

Considering the whole data set, the relative standard deviation is close to the maximum values depicted in Figure 4 (in the case of large and intermediate ions close to that of daytime, and in the case of small ions, close to that of nighttime values). The crossing point of the curves for daytime and nighttime, at a mobility of about 0.5 cm^2 V^{-1} s^{-1} (1.6 nm) in Figure 4, is in accordance with the boundary between cluster ions and aerosol ions [*Tammet*, 1995]. The above estimates of the relative standard deviations are equally valid for the large and intermediate air ion concentrations of both polarities, taking into account the random measuring errors.

The estimates of relative standard deviations of fraction concentrations of air ions in the region of large ions show quite a good agreement with those of aerosol measurements in a diameter interval of 10–100 nm [*Kikas et al.*, 1996]. According to the latter measurements, the relative standard deviation of aerosol particle concentrations has a minimum value in the size range of the accumulation mode, 100–300 nm, and rises in the flanks. This is also in accord with model calculations of deposition velocities of aerosol particles [*Jaenicke*, 1982, 1984; *Hoppel et al.*, 1990].

Normally, the positive air ion spectrum has a mode in a mobility range of 1.0–1.3 cm^2 V^{-1} s^{-1} or 1.3–1.6 cm^2 V^{-1} s^{-1}, and the negative ion spectrum, in a mobility range of 1.3–1.6 cm^2 V^{-1} s^{-1} or 1.6–2.0 cm^2 V^{-1} s^{-1}. Sometimes the "low mobility mode" of 1.0–1.3 cm^2 V^{-1} s^{-1} becomes dominant in the negative ion spectrum, and the mobility spectrum of negative small ions expands over a wider region as compared to positive ions. The mode of positive ion spectrum only shifts from mobilities of 1.3–1.6 cm^2 V^{-1} s^{-1} to 1.0–1.3 cm^2 V^{-1} s^{-1}. These variations could explain the higher relative standard deviation of the big cluster ion concentration of negative polarity compared with the ions of positive polarity.

The low-mobility modes of small air ions of both polarities were recorded when the large ion concentration was decreasing, but not vice versa. The low concentration of heavy large ions allows small air ions to evolve (grow) toward clusters of large sizes, and consequently to lower mobilities, within their lifetime. The evolution of the mobility spectra of small air ions described above was more regular in the warm season under conditions of anticyclones, particularly in June and August. In June and August, under conditions of hot and sunny anticyclonic weather, the low-mobility mode of negative ions in a mobility range of 1.0–1.3 cm^2 V^{-1} s^{-1} preferentially became dominant in the afternoon (or in the evening) and disappeared before sunset.

Considering the whole data set, the fractions of small ions of mobilities of 1.0–1.3 cm^2 V^{-1} s^{-1} and 1.3–1.6 cm^2 V^{-1} s^{-1} are the most conservative; the fractions of higher or lower mobility show higher relative standard deviations. Accordingly, the mobility of 1.3 cm^2 V^{-1} s^{-1} (diameter of 0.88 nm) may be used as a conventional boundary between small and big cluster ions. The fractions of negative small cluster ions at daytime display almost equal variabilities.

In general, small air ion concentrations have higher relative standard deviations in the nighttime than in the daytime, because of higher concentrations raised during nocturnal calms, in fine weather conditions in the warm season. The highest relative standard deviations of small cluster ion concentration were recorded in July in conditions of very hot and stable anticyclones, probably due to increasing ionization rate caused by accumulation of radon, thoron, and their daughters near the ground during nocturnal calms that produced numerous new young ions. The higher the mobility of small cluster ions, the higher the relative standard deviation was: for example, 40% for the ions of 1.3–1.6 cm^2 V^{-1} s^{-1} and 60% for the ions of 2.5–3.14 cm^2 V^{-1} s^{-1}.

3.2. Average Characteristics and Variability of Main Ion Groups

3.2.1. Small ions. The statistical characteristics of small air ion concentrations are presented in Table 2. The average concentrations of small air ions and their standard deviations are $n_- = 245 \pm 88$ cm^{-3} and $n_+ = 274 \pm 96$ cm^{-3}. The correlation coefficient between the polar concentrations is 98%. The mean natural mobility of small air ions of both polarities is calculated by averaging over the mobility interval from 0.5 to 3.2 cm^2 V^{-1} s^{-1}. The hourly mean mobilities and standard deviations, averaged over the whole measurement period of 14 months, are $k_- = 1.53 \pm 0.10$ and $k_+ = 1.36 \pm 0.06$ cm^2 V^{-1} s^{-1}. Approximately the same values of mean mobility have been found for different annual periods from 1985 to 1994. The correlation coefficient between polar mean mobilities is 80%. The mean mobility of small ions reduced to standard conditions is not discussed here because of the complicated nonlinear character of the reduction procedure [*Tammet*, 1998].

The frequency distributions of the concentration of positive small ion categories (original fractions P_1–P_8, classes of small and big clusters and their total concentration) are approximately lognormal and can be derived from the moments of distribution presented in Table 2. In the case of positive small

Table 2. Statistics of Negative/Positive Small Ion Concentrations (cm^{-3})

Mobility, cm^2 V^{-1} s^{-1}	Mean	Median	Maximum	Lower Quartile	Upper Quartile	Relative s.d.
2.51–3.14	12/7	10/7	67/42	8/5	13/8	0.45/0.51
2.01–2.51	33/19	29/17	180/99	24/14	37/23	0.41/0.44
1.60–2.01	56/45	51/41	265/207	44/34	62/51	0.37/0.38
1.28–1.60	59/69	55/64	252/303	44/54	68/79	0.36/0.35
1.02–1.28	42/69	40/66	157/284	28/54	54/83	0.44/0.36
0.79–1.02	24/41	21/39	107/154	12/29	33/51	0.59/0.40
0.63–0.79	13/16	11/14	83/74	6/10	18/21	0.68/0.50
0.50–0.63	8/8	6/7	58/45	4/5	9/10	0.72/0.59
0.50–3.14	245/274	231/259	990/1167	183/210	290/319	0.36/0.35
1.28/1.00–3.14	159/209	148/196	737/928	124/162	178/238	0.37/0.35
0.50–1.28/1.00	86/65	78/61	361/239	50/44	115/82	0.51/0.43

Number of measurements: 8615.

cluster ions, the distribution of the largest extreme gives a closer approximation, and in the case of big cluster ions, the gamma distribution is closer. The concentration of negative small cluster ions behaves similarly, but the concentration of negative big cluster ions shows different character. Its frequency distribution is extremely asymmetric, with a maximum at about 45 cm^{-3}, below the lower quartile (see Table 2).

The observed average values of mean mobility are comparable with those found by *Dhanorkar and Kamra* [1992], 1.37 and 1.25 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ for negative and positive polarity, respectively. The average values of reduced mobility at STP, as reported by *Mohnen* [1977], are 1.24 and 1.14 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ for negative and positive polarity, respectively. In both cases the ratio of negative to positive mobility is about 1.1.

The mean natural mobility of small air ions is higher in winter than in summer. The averaged mean mobility values and their standard deviations recorded in December and May are $k_{-\mathrm{Dec}} = 1.63 \pm 0.09\ \mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$, $k_{+\mathrm{Dec}} = 1.41 \pm 0.04\ \mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ and $k_{-\mathrm{May}} = 1.47 \pm 0.09\ \mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$, $k_{+\mathrm{May}} = 1.32 \pm 0.05\ \mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$. An analogous difference of mobilities was also found formerly [*Hõrrak et al.*, 1994].

The average negative and positive polar conductivities are nearly equal; that can be explained by the relatively high position (5 m) of air inlet and the screening of the electric field by trees surrounding the building where the instrumentation is located. The average polar conductivities calculated according to the entire measured mobility interval of 0.00032–3.2 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ are $\lambda_- = 6.18 \pm 2.14$ fS m^{-1} and $\lambda_+ = 6.18 \pm 2.14$ fS m^{-1}. These polar conductivities are nearly equal with those of small ions $\lambda_{-s} = 5.96 \pm 2.11$ fS m^{-1} and $\lambda_{+s} = 5.97 \pm 2.11$ fS m^{-1}; the increased average mobility of negative small ions as compared to that of positive ions entirely compensates differences in the concentration, on average. The ratio of positive ion concentration to that of negative ions (coefficient of unipolarity) is 1.127 ± 0.074, and the ratio of the average mobility of negative ions to that of positive ions is 1.124 ± 0.049. A regression analysis shows that the polar total conductivities are nearly equal; considering the entire range of measured values of conductivities from about 1.5 to 26 fS m^{-1}, the correlation coefficient is 99%. The conductivity (also small ion concentration) underwent decrease since 1985–1986 from about 9 fS m^{-1} to 6 fS m^{-1} in 1993–1994.

In fine weather conditions, both the mean mobility and the total concentration of small ions have the average diurnal variation of a single wave shape with a maximum in the nighttime and a minimum in the afternoon [*Hõrrak et al.*, 1998b]. The concentration of small ions has some considerable diurnal variation only in the warm season when the soil is unfrozen. The average diurnal variation of negative small ions is caused mainly by small cluster ions.

The absolute maximum of the total concentration of small ions recorded on August 26 was 996 cm^{-3} for negative ions and 1176 cm^{-3} for positive ions, both in early morning hours (0830 LST) in fine weather conditions. The origin of the high concentrations was probably the accumulation of radon and thoron near the ground during nocturnal calms. Daytime minimum values were 239 cm^{-3} for negative ions and 269 cm^{-3} for positive ions. The absolute maxima were recorded at the end of a 3-day period of very weak winds at daytime and calm in the nighttime. The highest concentration of the high-mobility fraction of small cluster ions (2.51–3.14 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$) was also recorded in the same morning at 0830 LST, when the negative ion concentration was 65 cm^{-3} and the positive ion concentration was 38 cm^{-3}. The minimum values recorded on the same day at the afternoon were about 10 cm^{-3} and 5 cm^{-3} for negative ions and positive ions, respectively. This means that during nighttime and early morning hours there exists some amount of very young ions with mobilities higher than 3.14 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ generated by the radioactivity of radon and thoron and their daughters. These ions remain out of scope of the mobility spectrometers. At other times their concentration is comparable with the measurement uncertainties. The largest contribution of the fraction (2.51–3.14 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$) to the total concentration of polar small ions was 7% and 4% for negative ions and positive ions, respectively. The estimated amount of high-mobility cluster ions (higher than 3.14 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$) during nighttime and early morning hours was about 2 times less than the above mentioned 7% and 4% from the total polar concentration of small ions for negative ions and positive ions, respectively.

3.2.2. Intermediate ions. The statistical characteristics of intermediate ion concentrations are presented in Table 3. The average concentration of intermediate ions is relatively low, about 50 cm^{-3}, but occasionally very high concentrations are recorded owing to the bursts of intermediate ions with concentrations of up to 900 cm^{-3} [*Hõrrak et al.*, 1998b]. The bursts occurred around local noon, the enhanced concentrations were recorded from 1000 to 1900 LST, with a duration of 6–10 hours, in fine weather conditions. The correlation coefficient between the total concentrations of positive and negative intermediate ions is 97%.

A burst of intermediate ions can initiate a process of the evolution of aerosol ions generating new aerosol particles that grow toward large sizes. This process looks like a triggering of a nucleation process with the accumulation of particles in the nucleation mode size range of 9.7–15 nm. Besides the process of evolution, a process of another character was also observed: a spectral mode suddenly appeared in the nucleation size range of 9.7–15 nm (mobility range of 0.0091–0.021 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$) or 15–22 nm (0.0042–0.091 $\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$)

Table 3. Statistics of the Negative/Positive Intermediate Ion Concentrations (cm^{-3})

Mobility, $\mathrm{cm}^2\,\mathrm{V}^{-1}\mathrm{s}^{-1}$	Mean	Median	Maximum	Lower Quartile	Upper Quartile	Relative s.d.
0.40–0.50	5/5	4/5	49/48	3/3	6/6	0.82/0.73
0.32–0.40	3/4	2/3	52/37	1/2	4/5	1.15/0.80
0.25–0.32	3/2	2/2	56/42	1/1	3/3	1.21/1.10
0.150–0.293	8/7	5/5	155/116	4/4	8/7	1.19/1.08
0.074–0.150	12/12	8/8	279/250	6/6	12/12	1.28/1.23
0.034–0.074	25/25	18/18	447/437	14/14	25/25	1.14/1.16
0.034–0.50	57/55	40/41	1008/874	31/32	58/57	1.08/1.03
0.25–0.50	11/12	9/10	157/116	6/7	13/13	0.95/0.78
0.034–0.293	45/44	31/31	851/761	24/24	45/44	1.14/1.13

Table 4. Number of Days in the Month When the Concentration of Positive Intermediate Ions Exceeds a Certain Value

		>100 cm^{-3}	>200 cm^{-3}	>300 cm^{-3}	>400 cm^{-3}	>500 cm^{-3}	>600 cm^{-3}	Maximum cm^{-3}
Sept.	1993	13	8	3	0	0	0	381
Oct.	1993	8	8	7	5	1	1*	874
Nov.	1993	5	3	3	0	0	0	442
Dec.	1993	2	1	1	1*	0	0	457
Jan.	1994	1	1	1	1*	0	0	471
Feb.	1994	2	2	2	2	2	1	601
March	1994	8	6	3	2	1	1	708
April	1994	5	2	2	2	1	1	860
May	1994	18	12	7	6	4	1	712
June	1994	11	6	2	1	0	0	470
July	1994	5	1*	0	0	0	0	205
Aug.	1994	6	1	0	0	0	0	205
Sept.	1994	6	2	2	1	1	0	506
Oct.	1994	11	6	3	2	1	1	652
	Sum	101	59	36	23	11	6	

* Burst of short duration: October and December within 2 hours, January within 3 hours more than 100 cm^{-3}.

and remained there for 4–8 hours (during the time of intensive sunlight), slightly changing in height. Several such events have been observed when an anticyclonic air mass of good visibility has come over the Baltic Sea to inland areas. In general the disturbed region of air ion mobility spectra affected by the bursts of intermediate ions was observed from about 0.002 to 1.0 $cm^2 V^{-1} s^{-1}$ (from 1.1 to 34 nm), including the groups of big cluster ions and light large ions.

In contrast to the light intermediate ions (0.32–0.50 $cm^2 V^{-1} s^{-1}$), the fraction concentrations of which are nearly equal, the heavy intermediate ions (0.034–0.293 $cm^2 V^{-1} s^{-1}$) show a rise in concentration toward lower mobilities. During the days with intermediate ion bursts, the ratio of concentrations of heavy to light fraction varies from about 3 to 7. The frequency distributions of the concentration of light intermediate ion fractions are asymmetric, approximately lognormal, because of the burst events. The concentrations of heavy intermediate ion fractions are roughly lognormally distributed because of extremely high values recorded during the burst events.

In order to present some statistical description of the bursts of intermediate ions, the number of days in the month when the concentration of intermediate ions is higher than a certain value is given in Table 4. Only more pronounced burst events are considered, when the intermediate ion polar concentration exceeds 100 cm^{-3} (background of about 50 cm^{-3}) during at least 2 hours. Commonly, the burst duration was 6–10 hours (from background up to maximum and down to background). The bursts of shorter duration are marked by an asterisk in Table 4. The same is also true for intermediate ions of negative polarity, but sometimes during the burst events, peak values of the concentration of negative intermediate ions exceed those of positive polarity by about 100–150 cm^{-3}.

In the period from November 24 to February 24, only three bursts were recorded, and even those bursts were of short duration, being higher than 100 cm^{-3} for only 2–3 hours. This is probably due to the fact that the conditions in winter did not favor photochemical nucleation because of low solar radiation intensity and duration at this latitude. There may also be a low concentration of nucleating low-pressure vapors. Also the decreasing mixing rate of the boundary layer and the accumulation of aerosol pollutants may be responsible for the absence of burst events in wintertime.

Regular bursts started as early as February 25 and 26, when bursts up to 600 cm^{-3} were recorded. The higher concentrations in May could be related to the beginning of the period of early vegetation and intensive agricultural works. On the basis of the side-by-side measurements of aerosol particle size spectra [*Hõrrak et al.*, 1996, 1998a] it can be concluded that the period of intensive bursts of intermediate ions followed the inflow of cool and clean high-pressure air mass. This inflow occurred on May 1, when the concentration of particles in the accumulation size range (100–560 nm) decreased rapidly from about 2400 cm^{-3} to 100 cm^{-3} and, after that, started gradually to increase again.

J. Mäkelä (personal communication, 1998) and *Birmili* [1998] have found the same regularities: the low concentration of ultrafine aerosol particles below 10 nm during wintertime and bursts in spring. The number of days with nucleation events of 3–5 nm particles observed at *Hyytiälä* forest station, southern Finland, was 56 during a 1-year period in 1996–1997 (J. Mäkelä, personal communication, 1998). This number has the same order of magnitude as the number of bursts of intermediate ions, about 80, for a 1-year period at Tahkuse in 1993–1994. The number of nucleation events found by *Birmili* [1998] in central Europe, near Leipzig, was between 38 and 60 during different seasons in 1996–1997.

3.2.3. Large ions. The statistical characteristics of large ion concentrations are presented in Table 5. As compared with intermediate ions, the frequency distributions of the concentration of light large ion categories (original fractions 15–17 and their total concentration) are closer to lognormal. As an exception, the frequency distribution of the fifteenth fraction shows similarity to that of intermediate ions. In all cases the frequency distributions are asymmetric because of high outliers. The frequency distributions of the concentration of heavy large ion categories (original fractions 18–20 and total concentration) are close to gamma distribution. There is no substantial difference between large ions of negative and positive polarity.

The whole range of large ions 0.00041–0.034 $cm^2 V^{-1} s^{-1}$ (diameters of 7.4–79 nm) can be divided into two classes with mobilities of 0.0042–0.034 $cm^2 V^{-1} s^{-1}$ (7.4–22 nm) and 0.00041–0.0042 $cm^2 V^{-1} s^{-1}$ (22–79 nm) called, by convention, light large ions and heavy large ions, respectively. In general, the ratio of concentrations of light large and heavy large ions is low, about 0.2, but in some cases (nucleation events) the ratio may be extremely high, up to about 2.5. These two categories show different behavior in the case of bursts of intermediate ions, when enhanced concentrations of light large ions have also been recorded. As a rule, the concentration of heavy large ions decreases before the burst of intermediate ions [*Hõrrak et al.*, 1998b].

Examining the time series of heavy large ion concentration, it was found that besides short time variations (bursts with duration of less than 1 day), this fraction also has a variation of 4–6 days (typical synoptical period) and even long time trends (1–2 weeks or more). The short time variations have higher amplitudes of 2000–4000 cm^{-3}, that is, about 10 times higher than the amplitude of average diurnal variation. The average diurnal variation is weak, about 150 cm^{-3}, with a minimum in the afternoon at 1300–1400 LST. The short time variations

Table 5. Statistics of the Negative/Positive Large Ion Concentration (cm^{-3})

Mobility, $cm^2V^{-1}s^{-1}$	Mean	Median	Maximum	Lower Quartile	Upper Quartile	Relative s.d.
0.0016–0.034	42/45	30/34	806/810	23/25	44/46	1.07/1.03
0.0091–0.0205	97/96	74/74	1673/1684	53/53	107/107	0.93/0.97
0.0042–0.0091	162/157	139/128	1874/1852	96/91	187/182	0.73/0.79
0.00192–0.0042	282/282	252/251	2139/2172	177/171	343/344	0.58/0.60
0.00087–0.00192	463/493	433/460	2604/2626	294/321	595/627	0.50/0.47
0.00041–0.00087	609/610	567/562	3119/3311	380/380	787/792	0.51/0.50
0.00041–0.034	1655/1783	1540/1578	8057/8099	1097/1116	2068/2103	0.48/0.47
0.0042–0.034	301/297	245/238	4276/4279	179/168	339/336	0.80/0.84
0.00041–0.0042	1354/1385	1258/1295	6881/6908	883/910	1720/1751	0.47/0.47

are probably caused by pollutant transport processes. Considering cold and warm seasons separately, the average diurnal variations display different behavior. In the cold season, minimum concentrations of about 1000 cm^{-3} were recorded in early morning hours at 0600–0700 LST, and maximum concentrations of about 1400 cm^{-3} were recorded in late evening at 2100 LST. In the warm season, minimum concentrations of about 1200 cm^{-3} were recorded in the afternoon at 1500 LST, and maximum concentrations of about 1400 cm^{-3} were recorded in early morning hours at 0700 LST.

Koutsenogii [1997] concluded that submicron particles with modal diameter of 170 nm represent regional and global aerosols. Such particles have the longest residence time of about 10 days and can be transported over distances of up to 8000 km. The heavy large ions as particles with diameters of 23–80 nm have residence times of 1.5–6 days [*Jaenicke*, 1982]; they are transported by air masses, are accumulated in the atmosphere, and are removed from the atmosphere preferentially by precipitation and thermal diffusion. The estimates of decay constants (similar to the residence times) given by *Hoppel et al.* [1990] for marine boundary layer are considerably smaller, about 10–30 hours, considering diameters of 20–80 nm. The latter estimates also take into account the loss of particles in cloud processes.

The average diurnal variation of the concentration of light large ions is similar to that of intermediate ions during the burst events, but their maximum is recorded with some time lag [*Hõrrak*, 1998b]. The time lag is explained by the growth of generated new particles toward larger sizes. At other times, their diurnal variation is weak and close to that of heavy large ions.

3.3. Statistical Classification of Air Ions According to Their Mobilities

3.3.1. Principal component and factor analysis. The principal component analysis (PCA), known in multivariate mathematical statistics, is applied to detect the structure of the air ion mobility spectrum, e.g., for the search of mobility boundaries between different groups of air ions. Fraction concentrations of a mobility spectrum of air ions may be interpreted as a set of variables that are closely correlated (see Table 6). The formal correlation is caused by the following: (1) physical and chemical processes embracing a group of fractions (causing positive correlation) or acting between different groups of fractions (causing opposite correlation) and (2) unavoidable smoothing of a spectrum due to the finite resolution of the measuring apparatus. The information about variance and covariance, which is included in different fractions of a mobility spectrum, can be transferred by a considerably smaller number of new variables, called principal components or factors, which are proper linear combinations of original variables. The search for principal components reduces to a

Table 6. Correlation Coefficients (in Percent) Between Negative Air Ion Mobility Fractions, September 1, 1993, to October 27, 1994

	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}	N_{11}	N_{12}	N_{13}	N_{14}	N_{15}	N_{16}	N_{17}	N_{18}	N_{19}	N_{20}
N_1	100	85	76	62	38	24	19	13	12	3	13	12	7	2	-4	3	4	-5	-8	-8
N_2	85	100	93	80	51	31	23	11	6	-4	9	8	1	-8	-16	-11	-11	-18	-22	-24
N_3	76	93	100	92	65	41	28	15	5	-4	5	4	-4	-12	-19	-17	-17	-21	-29	-35
N_4	62	80	92	100	88	70	56	36	22	6	12	12	1	-7	-14	-11	-14	-22	-36	-48
N_5	38	51	65	88	100	94	82	61	41	21	24	24	13	6	-2	1	-2	-17	-35	-55
N_6	24	31	41	70	94	100	93	73	51	28	32	32	20	13	5	9	5	-13	-32	-52
N_7	19	23	28	56	82	93	100	82	63	40	44	44	33	26	17	20	15	-4	-23	-42
N_8	13	11	15	36	61	73	82	100	72	54	56	57	48	42	34	35	30	13	-5	-23
N_9	12	6	5	22	41	51	63	72	100	66	66	68	63	59	52	50	44	29	12	-5
N_{10}	3	-4	-4	6	21	28	40	54	66	100	69	71	66	60	52	46	40	31	18	5
N_{11}	13	9	5	12	24	32	44	56	66	69	100	97	87	73	58	43	34	22	10	1
N_{12}	12	8	4	12	24	32	44	57	68	71	97	100	89	75	58	43	33	20	9	0
N_{13}	7	1	-4	1	13	20	33	48	63	66	87	89	100	92	78	61	47	34	19	6
N_{14}	2	-8	-12	-7	6	13	26	42	59	60	73	75	92	100	93	77	62	47	29	13
N_{15}	-4	-16	-19	-14	-2	5	17	34	52	52	58	58	78	93	100	88	77	65	44	23
N_{16}	3	-11	-17	-11	1	9	20	35	50	46	43	43	61	77	88	100	92	78	54	31
N_{17}	4	-11	-17	-14	-2	5	15	30	44	40	34	33	47	62	77	92	100	90	70	46
N_{18}	-5	-18	-21	-22	-17	-13	-4	13	29	31	22	20	34	47	65	78	90	100	89	66
N_{19}	-8	-22	-29	-36	-35	-32	-23	-5	12	18	10	9	19	29	44	54	70	89	100	87
N_{20}	-8	-24	-35	-48	-55	-52	-42	-23	-5	5	1	0	6	13	23	31	46	66	87	100

The absolute value of critical correlation coefficient at a confidence level of 95% is 3%.

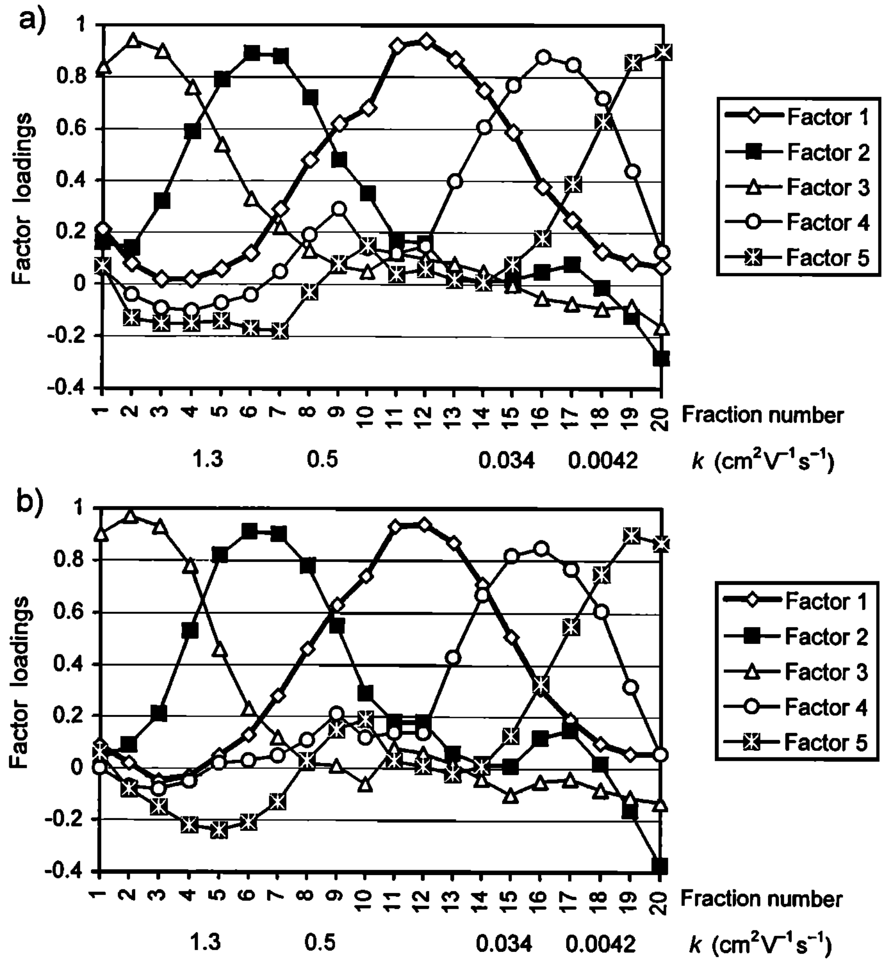


Figure 5. Factors of air ion mobility spectra for (a) positive ions and (b) negative ions. The mobility and diameter boundaries of fractions are given in Table 1.

search for eigenvalues (characteristic roots, portions of common variance explained by factors) and factor loadings (characteristic vectors) of a correlation matrix of original variables.

Before performing PCA, the original variables (fractions of air ion mobility spectra) were treated with a nonlinear transformation by logarithmic scaling. This procedure transforms asymmetric frequency distributions of variables closer to the normal ones, assumed by PCA. The logarithmic scaling does not significantly affect the results of the classification of air ions in our case. Finally, the variables were standardized to provide variables of a comparable variance. To obtain a clear pattern ("*simple structure*") of loadings, the VARIMAX rotation, often used in factor analysis, has been performed herein.

The eigenvalue problem was solved separately for the correlation matrices of logarithmically rescaled and standardized variables of positive and negative ions (Table 6). There is some clear structure in these correlation matrices. Results are presented in Figure 5 for positive and negative ions, respectively. The boundaries of spectral fractions and corresponding diameter intervals for single charged particles are given in Table 1.

The first five successfully extracted factors explain 92% of total variance. The total variance that can be potentially extracted is equal to the number of variables, which is 20. Each of the first five factors extracts at least as much variance as the equivalent of one original variable, i.e., 5% (it is expected that the variance of a single standardized variable is 1); a deep drop follows thereafter. The subsequent 14 factors explain only 8% of the total variance. Each of the latter factors explains less than 1.5% of the total variance. A part of this variation is caused by instrumental noise. Thus we can conclude that the mobility spectrum, in the first approximation, has 5 degrees of freedom or that the spectrum can be described almost completely by these five factors representing 92% of all measured information.

The first factor (factor 1 in Figure 5) is closely correlated with intermediate ions (fractions 9–14), and thus it can be called the "burst factor" of intermediate ions. It explains 24% of variance, more than others do. Factor 2 is closely correlated with big cluster ions (fractions 4–8); factor 3, with small cluster ions (fractions 1–4); and factor 4, with light large ions (fractions 15–18). They explain approximately equal variances of 20%, 18%, and 17% for factors 2, 3, and 4, respectively. The contribution of factor 5, associated with heavy large ions (fractions 18–20), is the lowest, 13%. This factor also is correlated oppositely with cluster ions (fractions 2–7). In the same sense, factor 2, which is closely correlated with big cluster ions (fractions 5–8), is correlated negatively with heavy large ions (fractions 19–20).

3.3.2. Air ion classes. The study of the correlation between the factors and air ion fractions shows that all the air ions can be divided into two main classes: (1) aerosol ions

Table 7. Classification of Air Ions

Class of Air Ions	Mobility, cm^2 V^{-1} s^{-1}	Diameter, nm	Traditional Name
Small cluster ions	1.3–3.2	0.36–0.85	small ions
Big cluster ions	0.5–1.3	0.85–1.6	small ions
Intermediate ions	0.034–0.5	1.6–7.4	intermediate ions
Light large ions	0.0042–0.034	7.4–22	large ions
Heavy large ions	0.00041–0.0042	22–79	large ions

with mobilities below 0.5 cm^2 V^{-1} s^{-1} and (2) cluster ions with mobilities above 0.5 cm^2 V^{-1} s^{-1}. These two classes can in turn be divided into two classes of cluster ions (small and big cluster ions) and three classes of aerosol ions (intermediate, light, and heavy large ions). The classification, based on statistical analysis, is given in Table 7. This classification is still to a certain extent conventional, and the boundaries are not exactly determined, because the factors that were chosen as representative have cross loadings (any variable is correlated with more than one factor; see Figure 5).

Considering the warm season (from May to September) separately from the entire period, the factor analysis revealed different boundaries between small cluster ions and big cluster ions of different polarity: negative small and big cluster ions have a boundary of 1.3 cm^2 V^{-1} s^{-1}, and positive ions, 1.0 cm^2 V^{-1} s^{-1} (diameter of 1 nm). The mobility boundary of 1.3 cm^2 V^{-1} s^{-1} halves the peak in the mobility spectrum of positive small ions (see Figure 3). If we use the boundary of 1.3 cm^2 V^{-1} s^{-1} for cluster ions of both polarities, then we obtain a lower concentration of positive small cluster ions compared to negative ions, and this would be in contradiction to our understanding about the electrode effect near the ground. The use of the boundary of 1.0 cm^2 V^{-1} s^{-1} for positive ions also facilitates the description of the average diurnal variation of cluster ion characteristics. Therefore we suggest the use of a boundary of 1.0 cm^2 V^{-1} s^{-1} between small and big cluster ions of positive polarity instead of 1.3 cm^2 V^{-1} s^{-1}. However, such a specification is rather speculative; measurements of small ion mobility spectra with higher resolution are necessary to establish the boundary more precisely.

In the warm season, the boundary between light and heavy large ions is shifted to a lower mobility of 0.00192 cm^2 V^{-1} s^{-1} (diameter of 34 nm) compared to that of the cold season 0.0043 cm^2 V^{-1} s^{-1} (diameter of 22 nm).

The boundary mobility of 0.5 cm^2 V^{-1} s^{-1} or a diameter of 1.6 nm is the same boundary, which has been considered physically as the boundary between molecular clusters and macroscopic particles [*Tammet*, 1995]. The same value of 0.5 cm^2 V^{-1} s^{-1} was also considered as the lower boundary of small air ions formerly [*Hõrrak et al.*, 1994].

The classification of air ions presented in Table 7 may also be obtained by PCA without the Varimax rotation procedure, using the first two factors (with respect to eigenvalues) as classifiers. In this case, at first, the boundary between cluster ions and aerosol ions can be determined more accurately when excluding the burst events of intermediate ions. The subsequent classification within separated classes of cluster ions and aerosol ions makes it possible to gradually detail the boundaries between different classes of air ions. The presented classification is in general also predictable from the average spectrum and from the relative standard deviations of fraction concentrations (see Figures 2, 3, and 4).

The above classes of air ions could be physically characterized as follows:

1. Small cluster ions have mobility of 1.3–2.5 cm^2 V^{-1} s^{-1}, estimated diameter of 0.36–0.85 nm, mass of 30–400 unified atomic mass units (u), and typical lifetime of 5–60 s. Considering ion diameters, the core of a cluster could contain one inorganic molecule and be surrounded by one layer of water molecules. After recombination, small cluster ions would be destroyed and again separated into initial components (cores and water molecules).
2. Big cluster ions have mobility of 0.5–1.3 cm^2 V^{-1} s^{-1}, estimated diameter of 0.85–1.6 nm, and mass of 400–2500 u. Considering ion diameters, the core of a cluster could contain one organic molecule and be surrounded by a layer of water molecules. The enhanced concentrations have been recorded when large ion concentration is low, which makes it possible for them to evolve to large sizes within their longer lifetime. In the case of intensive nucleation events (bursts), the enhanced concentrations were recorded simultaneously with intermediate ion concentrations. Contrary to aerosol ions, collisions between cluster ions and ambient gas molecules are considered elastic [*Tammet*, 1995].
3. Intermediate ions have mobility of 0.034–0.5 cm^2 V^{-1} s^{-1} and diameter of 1.6–7.4 nm. A corresponding class of aerosol particles is *the "fine nanometer particles"*. Some intermediate ions are a product of ion-induced nucleation: nucleating vapor condenses onto cluster ions, which grow to the size of intermediate ions, called *the "primary aerosol ions"*. Particles born in the neutral stage in the process of gas-to-particle conversion or nucleation and charged by attachment of cluster ions are called *the "secondary aerosol ions"*.
4. Light large ions have mobility of 0.0042–0.034 cm^2 V^{-1} s^{-1} and diameter of 7.4–22 nm. A corresponding class of aerosol particles is *the "ultrafine particles"* or *"coarse nanometer particles"*. They are single charged and often in a quasi-steady state of stochastic charging with cluster ions.
5. Heavy large ions have mobility of <0.0042 cm^2 V^{-1} s^{-1} and diameter of >22 nm. A corresponding class of aerosol particles could be called *the "Aitken particles"*. They are, as a rule, in a quasi-steady state of stochastic charging with cluster ions, and some of them may carry multiple charges.

We suppose that small cluster ions represent a group of young ions and big clusters represent a group of aged ions. This assumption is in accordance with the measurements of the mobility spectra of ions generated in laboratory air [*Nagato and Ogawa*, 1998]. They have found no ions below 0.8 cm^2 V^{-1} s^{-1} in the mobility spectrum of young ions, while a considerable number of ions was observed down to 0.3 cm^2 V^{-1} s^{-1} in the spectrum of natural ions. It was supposed that the cluster ions between 0.3 and 0.8 cm^2 V^{-1} s^{-1} could be formed by mechanisms other than those for the ions above 0.8 cm^2 V^{-1} s^{-1}. Our measurements show the boundary between two groups at 1.0 cm^2 V^{-1} s^{-1} and 1.3 cm^2 V^{-1} s^{-1} for the ions of positive and negative polarity, respectively.

The presented classification of aerosol ions is in accord with the three-modal structure of submicron aerosol particle size distribution found in continental sites and in the Arctic marine boundary layer [*Kulmala et al.*, 1996; *Mäkelä et al.*, 1997; *Covert et al.*, 1996; *Birmili*, 1998]. These modes have mean diameters of about 150–250 nm, 40–70 nm, and 5–14 nm and are referred to as the accumulation, Aitken, and nucleation (or ultrafine) modes, respectively. There were clear minima in number concentrations between these modes that appeared at 20–30 nm and 80–100 nm. Thus the intermediate ions (0.034–0.5 $cm^2 V^{-1} s^{-1}$, 1.6–7.4 nm) and light large ions (0.0042–0.034 $cm^2 V^{-1} s^{-1}$, 7.4–22 nm) may be classified as two classes of nucleation mode particles, and heavy large ions (0.00041–0.0042 $cm^2 V^{-1} s^{-1}$, 22–79 nm), as charged Aitken mode particles. It may be concluded that in the atmosphere, there exists a natural boundary dividing ultrafine particles at about 7.4 nm, and when studying aerosol processes, the size range of 1.6–7.4 nm can be considered the range of fine nanometer particles.

4. Conclusions

The average spectrum of air ions in a wide mobility range of 0.00041–3.2 $cm^2 V^{-1} s^{-1}$ is established on the ground of a statistically weighty database: 14 months, 8615 hourly averaged spectra of both polarities, and 20 logarithmically distributed fractions per spectrum. The average spectrum gives a basis for distinguishing three main air ion groups: small, intermediate, and large ions. The groups of small and large ions are distinctly seen as peaks in an average mobility spectrum. Physically, large and intermediate ions can be called aerosol ions and small ions can be called cluster ions.

Small (cluster) ions represent quite isolated and stable groups of ions with mean natural mobility and standard deviation of 1.53 ± 0.10 and 1.36 ± 0.06 $cm^2 V^{-1} s^{-1}$ for negative and positive ions, respectively. The ratio of the concentration of positive ions to that of negative ions (coefficient of unipolarity) is about 1.12, as well as the ratio of the average mobility of negative ions to that of positive ions. Accordingly, the average polar conductivities and their standard deviations, caused mainly by small ions, are equal: $\lambda_- = 6.18 \pm 2.14$ fS m^{-1} and $\lambda_+ = 6.18 \pm 2.14$ fS m^{-1}. This can be explained by a relatively high position (5 m) of the air inlet and the screening of the electric field by trees surrounding the building where the instrumentation is located.

The overall shape of the spectra in the range of large ions (aerosol ions) is in accord with calculations based on the theory of bipolar charging of aerosol particles by small air ions. The concentration of large ions diminishes toward higher mobilities owing to a reduction of charging probability and of the concentration of aerosol particles.

Certain thermodynamical causes hinder the growth of cluster ions in ordinary environmental conditions, keeping the concentration of intermediate ions (0.034–0.5 $cm^2 V^{-1} s^{-1}$; 1.6–7.4 nm) at a low background of about 50 cm^{-3}. Enhanced concentrations of intermediate ions up to about 900 cm^{-3} (bursts) are observable in the mobility spectra in fine weather conditions during daytime. The number of burst events of intermediate ions recorded during 14 months was 101 (about 80 per year), with maximum frequency in spring and minimum frequency in winter. Intermediate ions are formed probably by diffusion charging of nanometer aerosol particles generated by photochemical nucleation process. At the same time, cluster ions can also grow up to intermediate ion size range. The measurements of intermediate ion mobility spectra may give essential information about nanometer particles and their electrical state in the case of nucleation events. These measurements may be used parallel to the nanometer particle measurements, in order to explain the importance of the possible routes of the generation of nanometer particles via homogeneous and ion-induced nucleation.

The relative standard deviation of the hourly averaged values of fraction concentration is about 50% for small (cluster) ions, 70% for large ions, and up to 130% for intermediate ions. The considerable variability of the concentration of intermediate ions is due to their bursts in favorable conditions during daytime. During nighttime, intermediate and large ions show nearly equal relative standard deviations of 50–60%.

The principal component analysis was applied to detect the structure of an air ion mobility spectrum, for example, for the search of mobility boundaries between different groups of air ions. The first five successfully extracted factors explain 92% of total variance. The study shows that a mobility spectrum in the range of 0.00041–3.2 $cm^2 V^{-1} s^{-1}$ (0.36–79 nm) can be divided into 5 classes: small cluster ions, big cluster ions, intermediate ions, light large ions, and heavy large ions with boundaries between them: 1.3 $cm^2 V^{-1} s^{-1}$ (diameter of 0.85 nm), 0.5 $cm^2 V^{-1} s^{-1}$ (1.6 nm), 0.034 $cm^2 V^{-1} s^{-1}$ (7.4 nm), and 0.0042 $cm^2 V^{-1} s^{-1}$ (22 nm). Thus it can be concluded that the mobility spectrum, in the first approximation, has 5 degrees of freedom or that the spectrum can be described almost completely by these five factors. The presented classification of aerosol ions is in accord with the three-modal structure of the size distribution of submicron aerosol particles.

Acknowledgments. This research has in part been supported by the Estonian Science Foundation through grants 3050, 3326, and 3903. The authors also thank Hilja Iher and, posthumously, Rein Sepp for their assistance in measuring.

References

Birmili, W., *Production of new ultrafine aerosol particles in continental air masses*, Ph.D. thesis, 107 pp., Univ. of Leipzig, Leipzig, Germany, 1998.

Covert, D. S., A. Wiedensohler, P. Aalto, J. Heintzenberg, P. H. McMurry, and C. Leck, Aerosol number size distributions from 3 to 500 nm diameter in the Arctic marine boundary layer during summer and autumn. *Tellus, 48*, Ser. B, 197–212, 1996.

Dhanorkar, S., and A. K. Kamra, Measurement of mobility spectrum and concentration of all atmospheric ions with a single apparatus, *J. Geophys. Res., 96*, 18,671–18,678, 1991.

Dhanorkar, S., and A. K. Kamra, Relation between electrical conductivity and small ions in the presence of intermediate and large ions in the lower atmosphere, *J. Geophys. Res., 97*, 20,345–20,360, 1992.

Dhanorkar, S., and A. K. Kamra, Diurnal variations of the mobility spectrum of ions and size distribution of fine aerosols in the atmosphere, *J. Geophys. Res., 98*, 2639–2650, 1993a.

Dhanorkar, S., and A. K. Kamra, Diurnal and seasonal variations of the small-, intermediate-, and large-ion concentrations and their contributions to polar conductivity, *J. Geophys. Res., 98*, 14,895–14,908, 1993b.

Flagan, R. C., History of electrical aerosol measurements, *Aerosol Sci. Technol., 28*, 301–380, 1998.

Hoppel, W. A., and G. M. Frick, Ion–aerosol attachment coefficients and the steady-state charge distribution on aerosols in a bipolar ion environment, *Aerosol Sci. Technol., 5*, 1–21, 1986.

Hoppel, W. A., J. W. Fitzgerald, G. M. Frick, R. E. Larson, and E. J. Mack, Aerosol size distributions and optical properties found in the marine boundary layer over the Atlantic Ocean, *J. Geophys. Res., 95*, 3659–3686, 1990.

Hõrrak, U., H. Tammet, J. Salm, and H. Iher, Diurnal and annual variations of atmospheric ionisation quantities in Tahkuse (in Russian with English summary), *Acta Commentat. Univ. Tartu, 824*, 78–83, 1988.

Hõrrak, U., F. Miller, A. Mirme, J. Salm, and H. Tammet, Air ion observatory at Tahkuse: Instrumentation, *Acta Commentat. Univ. Tartu, 880*, 33–43, 1990.

Hõrrak, U., H. Iher, A. Luts, J. Salm, and H. Tammet, Mobility spectrum of air ions at Tahkuse Observatory, *J. Geophys. Res., 99*, 10,697–10,700, 1994.

Hõrrak, U., J. Salm, and H. Tammet, Outbursts of intermediate ions in atmospheric air, in *Proceedings of the 10th Conference on Atmospheric Electricity*, pp. 76–79, Soc. of Atmos. Electr. of Japan, Osaka, 1996.

Hõrrak, U., A. Mirme, J. Salm, E. Tamm, and H. Tammet, Air ion measurements as a source of information about atmospheric aerosols, *J. Atmos. Res., 46(3–4)*, 233–242, 1998a.

Hõrrak, U., J. Salm, and H. Tammet, Bursts of intermediate ions in atmospheric air, *J. Geophys. Res., 103*, 13,909–13,915, 1998b.

Hõrrak, U., A. Mirme, J. Salm, E. Tamm, and H. Tammet, Study of covariations of aerosol and air ion mobility spectra at Tahkuse, Estonia, *J. Aerosol Sci., 29*, S849–S850, 1998c.

Israël, H., *Atmospheric Electricity, Vol. 1*, Isr. Program for Sci. Transl., Jerusalem, 1970.

Jaenicke, R., Physical aspects of the atmospheric aerosol. in: *Chemistry of the Unpolluted and Polluted Troposphere*, edited by H. W. Georgii and W. Jaeschke, pp. 341–373, D. Reidel, Norwell, Mass., 1982.

Jaenicke, R., Our knowledge about the atmospheric aerosol, in *Proceedings of the 11th International Conference on Atmospheric Aerosols, Condensation and Ice Nuclei*, vol. 1, pp. 99–107, Hungarian Meteorol. Service, Budapest, 1984.

Kikas, Ü., A. Mirme, E. Tamm, and T. Raunemaa, Statistical characteristics of aerosol in Baltic Sea region, *J. Geophys. Res., 101*, 19,319–19,327, 1996.

Kim, T. O., M. Adachi, K. Okuyama, and J. H. Seinfeld, Experimental measurement of competitive ion-induced and binary homogeneous nucleation in $SO_2/H_2O/N_2$ mixtures, *Aerosol Sci. Technol., 26*, 527–543, 1997.

Kim, T. O., T. Ishida, M. Adachi, K. Okuyama, and J. H. Seinfeld, Nanometer–sized particle formation from $NH_3/SO_2/H_2O$/air mixtures by ionizing irradiation, *Aerosol Sci. Technol., 29*, 111–125, 1998.

Kojima, H., Relation between intermediate ions and meteorological factors. *Res. Lett. Atmos. Electr., 4*, 49–53, 1984.

Koutsenogii, P., Aerosol measurements in Siberia, *J. Atmos. Res., 44*, 167–173, 1997.

Kulmala, M., A. Laaksonen, P. Aalto, T. Vesala, and L. Pirjola, Formation, growth, and properties of atmospheric aerosol particles and cloud droplets, *Geophysica, 32(1–2)*, 217–233, 1996

Luts, A., *Mathematical simulation of the evolution or air ions*, Ph.D. thesis, 150 pp., Univ. of Tartu, Tartu, Estonia, 1995.

Luts, A., and J. Salm, Chemical composition of small atmospheric ions near the ground, *J. Geophys. Res., 99*, 10,781–10,785, 1994.

Mäkelä, J. M., P. Aalto, V. Jokinen, T. Pohja, A. Nissinen, S. Palmroth, T. Markkanen, K. Seitsonen, H. Lihavainen, and M. Kulmala, Observations of ultrafine aerosol particle formation and growth in boreal forest, *Geophys. Res. Lett., 24*, 1219–1222, 1997.

Misaki, M., Studies of the atmospheric ion spectrum (I), *Pap. Meteorol. Geophys., 12*, 247–260, 1961a.

Misaki, M., Studies of the atmospheric ion spectrum (II), *Pap. Meteorol. Geophys., 12*, 261-276, 1961b.

Misaki, M., Mobility spectrums of large ions in the New Mexico semidesert, *J. Geophys. Res., 69*, 3309–3318, 1964.

Misaki, M., Measurements of atmospheric electricity (in Japanese), *Kisho Kenkyu Noto, 130*, 105–118, 1976.

Misaki, M., M. Ohtagaki, and I. Kanazawa, Mobility spectrometry of the atmospheric ions in relation to atmospheric pollution, *Pure Appl. Geophys., 100*, 133–145, 1972.

Mohnen, V. A., Formation, nature and mobility of ions of atmospheric importance, in *Electrical Processes in Atmospheres*, edited by H. Dolezalek and R. Reiter, pp. 1–17, Dietrich Steinkopff Verlag, Darmstadt, Germany, 1977.

Nagato, K., and T. Ogawa, Evolution of tropospheric ions observed by an ion mobility spectrometer with a drift tube, *J. Geophys. Res., 103*, 13,917–13,925, 1998.

Reischl, G. P., J. M. Mäkelä, R. Karch, and J. Necid, Bipolar charging of ultrafine particles in the size range below 10 nm, *J. Aerosol Sci., 27*, 931–949, 1996.

Salm, J., The average mobility spectrum of large ions of the troposphere, *Res. Lett. Atmos. Electr., 8*, 21–24, 1988.

Tammet, H., *The Aspiration Method for the Determination of Atmospheric-Ion Spectra*, Isr. Program for Sci. Transl., Jerusalem. 1970.

Tammet, H., A piecewise linear model of spectrum for the measurement of air ions and aerosols (in Russian with English summary), *Acta Commentat. Univ. Tartu, 534*, 45–54, 1980.

Tammet, H., Air ion observatory at Tahkuse: Software, *Acta Commentat. Univ. Tartu, 880*, 44–51, 1990.

Tammet, H., Size and mobility of nanometer particles, clusters and ions, *J. Aerosol Sci., 26*, 459–475, 1995.

Tammet, H., Reduction of air ion mobility to standard conditions, *J. Geophys. Res., 103*, 13,933–13,937, 1998.

Tammet, H., F. Miller, E. Tamm, T. Bernotas, A. Mirme, and J. Salm, Apparatus and methods for the spectrometry of small air ions (in Russian with English summary), *Acta Commentat. Univ. Tartu, 755*, 18–28, 1987.

U. Hõrrak, J. Salm, and H. Tammet, Institute of Environmental Physics, University of Tartu, 18 Ülikooli Street, Tartu 51014, Estonia. (Urmas.Horrak@ut.ee; Jaan.Salm@ut.ee; ht@ut.ee)

(Received July 9, 1999; revised November 24,1999;
accepted November 30, 1999.)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9303-9320, APRIL 16, 2000

Perturbation dynamics in atmospheric chemistry

Brian F. Farrell

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts

Petros J. Ioannou

Department of Physics, National and Capodistrian University of Athens, Athens

Abstract. Current understanding of how chemical sources and sinks in the atmosphere interact with the physical processes of advection and diffusion to produce local and global distributions of constituents is based primarily on analysis of chemical models. One example of an application of chemical models which has important implications for global change is to the problem of determining sensitivity of chemical equilibria to changes in natural and anthropogenic sources. This sensitivity to perturbation is often summarized by quantities such as a mean lifetime of a chemical species estimated from reservoir turnover time or the decay rate of the least damped normal mode of the species obtained from eigenanalysis of the linear perturbation equations. However, the decay rate of the least damped normal mode or a mean lifetime does not comprehensively reveal the response of a system to perturbation. In this work, sensitivity to perturbations of chemical equilibria is assessed in a comprehensive manner through analysis of the system propagator. When chemical perturbations are measured using the proper linear norms, it is found that the greatest disturbance to chemical equilibrium is achieved by introducing a single chemical species at a single location, and that this optimal perturbation can be easily found by a single integration of the transpose of the dynamical system. Among other results are determination of species distributions produced by impulsive, constant, and stochastic forcing; release sites producing the greatest and least perturbation in a chosen constituent at another chosen site; and a critical assessment of chemical lifetime measures. These results are general and apply to any perturbation chemical model, including three-dimensional global models, provided the perturbations are sufficiently small that the perturbation dynamics are linear.

1. Introduction

Models commonly used to assess the impact of perturbations to the chemical constituents of the atmosphere consist of continuity equations in which the time derivatives of the number densities of chemical species at grid points are related by terms representing advection, diffusion, sources, sinks, and reactions among the species. Given an initial state such a model can be integrated in time to obtain the species number densities as a function of the spatial and temporal distribution of sources. Although chemical equilibria are, in general, nonlinear, the response of the nonlinear equilibrium to sufficiently small perturbations is intrinsically linear, and even if the expected perturbations are large enough to require nonlinear terms in a comprehensive perturbation analysis, still a thorough understanding of the linear dynamics of small perturbations is a necessary foundation for this extension into the nonlinear regime. Therefore we assume in this work that the model is linear in the perturbations or can be linearized about a representative equilibrium state. For a description of such models, see the work of *Khalil and Rasmussen* [1984], *Chameides and Perdue* [1997], and *Prather* [1996, 1997, 1998]. Prather showed that the evolution of chemical perturbations to stationary equilibrium states can be obtained by decomposing the perturbation into the natural modes (the eigenvectors) of the linear system and that the least damped natural mode is the pattern that eventually dominates the response to any initial perturbation. He also noted that this asymptotic response is often masked by the presence of the other natural modes that are also excited by the initial perturbation and which interfere on superposition, leading to transient excursions of chemical per-

Copyright 2000 by the American Geophysical Union.

Paper number 1999JD901021.
0148-0227/00/1999JD901021$09.00

turbations, a phenomenon ascribed to the "nonorthogonality" of the modes. In light of the insights proceeding from application of the method of modes, identification of the natural modes in atmospheric chemical transport models has been advanced as the preferred method for describing the complex evolution of chemical perturbations [*Prather*, 1998].

However, systematic analysis of perturbation dynamics in chemical systems immediately encounters an impediment in making precise the concept of orthogonality among the natural modes. The familiar dot product allows both defining an angle between vectors as well as associating a length to each vector, and if chemical perturbations, described by the vector x, are measured by their euclidean length (otherwise called the L_2 norm), then the concept of orthogonality among chemical perturbations and among the normal modes of the system is definable and the transient growth of perturbations, in the L_2 sense, can be related to the orthogonality of the modes. A perturbation imposed on a stable system with orthogonal eigenvectors necessarily decays as measured in the L_2 norm at a rate faster than the decay rate of the least damped mode, and this limiting decay rate is approached asymptotically with time. (A system is stable if the real parts of its eigenvalues are negative.) For stable systems with nonorthogonal eigenvectors a perturbation measured in L_2 is not assured to decay at all times and may, in fact, increase for a period of time before eventually assuming its asymptotic decay rate, which is that of the least damped mode.

However suggestive these L_2 measures may be, chemical perturbations can only be sensibly assessed with a measure based on the number of molecules, and while such linear measures (as opposed to quadratic, e.g., L_2 measures) associate length to vectors, they do not allow definition of an angle between vectors. Nevertheless, while the concept of orthogonality is lost in these measures, the concept of projection of a perturbation onto eigenvectors remains valid, and the evolution of the perturbation still can be viewed in terms of summation over the evolving eigenvectors.

In order to understand the implications of this distinction between linear and quadratic measures, consider the abstract linear equation assumed to govern evolution of chemical constituents in a hypothetical chemical model:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} \ , \qquad (1)$$

where $\mathbf{A}$ is a real matrix and $\mathbf{x}$ is a vector containing as its entries the abundance of the chemical constituents in the model collocated at grid points. Associated with $\mathbf{A}$ is the transpose matrix $\mathbf{A}^T$. There is an intimate relation between the eigenvectors of these two matrices: the eigenvectors of $\mathbf{A}^T$ arranged in rows form the projection matrix that determines the coefficients of projection of a perturbation on the eigenvectors of $\mathbf{A}$. (If $\mathbf{U}$ are the eigenvectors of $\mathbf{A}$ arranged as columns, then the eigenvectors of $\mathbf{A}^T$ are $(\mathbf{U}^{-1})^T$. A vector $\mathbf{x}$ can be written as $\mathbf{x} = \mathbf{U}(\mathbf{U}^{-1}\mathbf{x})$, where $\mathbf{U}^{-1}\mathbf{x}$ is recognized as the coefficients of projection of the vector on the eigenvectors. The matrix $\mathbf{U}^{-1^T}$ is called the bi-orthogonal matrix. If $\mathbf{u}$ is a specific eigenvector of $\mathbf{A}$ corresponding to a given column $\mathbf{U}$, then its bi-orthogonal is the corresponding column of the bi-orthogonal matrix.) If the eigenvectors of $\mathbf{A}$ and $\mathbf{A}^T$ are the same, then the two matrices commute; that is, $\mathbf{A}\mathbf{A}^T - \mathbf{A}^T\mathbf{A} = 0$, and the system matrix $\mathbf{A}$ is called normal, otherwise nonnormal. In the chemical literature, where discretization of linear operators usually results in a real matrix, nonnormality of $\mathbf{A}$ can be expected when the matrix $\mathbf{A}$ exhibits asymmetry about the diagonal. (However, even for real matrices, nonnormality is not equivalent to asymmetry of the entries about the diagonal. Consider the orthogonal matrices which are asymmetric but nevertheless have the same bi-orthogonal and eigenvector matrices.)

Nonnormality of the system matrix $\mathbf{A}$ indicates that the eigenvalues and eigenvectors of the system may be illconditioned and highly sensitive to small perturbations in the parameters of the system [*Trefethen*, 1991]. As a result, projection on the natural modes may be illconditioned unless the parameters of the system are known with great certainty.

However, while nonnormality of the system identifies the potential for L_2 growth, accurate assessment of growth in other measures, such as the linear measures appropriate in chemistry, must proceed from detailed calculation in the measure selected. This can be done by defining carefully the measure of chemical perturbations and then determining the maximum chemical perturbation that can be achieved at any time in this measure. This task is facilitated by shifting attention from natural modes to the propagator (introduced in the next paragraph), which can reveal directly the perturbation dynamics of the system.

The temporal evolution of an arbitrary system of finite dimension can be expressed as

$$\mathbf{x}(t) = \mathbf{\Phi}(t)\,\mathbf{x}(0) \ , \qquad (2)$$

where $\mathbf{\Phi}(t)$ is the propagator, which in the case of a time-independent system is the matrix exponential $e^{\mathbf{A}t}$. (The matrix exponential $e^{\mathbf{A}t}$ is defined as $e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \mathbf{A}^2t^2/2! + \cdots$. There are efficient algorithms available for direct evaluation of the matrix exponential that do not require prior eigenanalysis of $\mathbf{A}$. We use a Padé approximation routine with scaling and squaring described in chapter 11.3 of *Golub and Van Loan* [1996].)

We wish to determine the initial perturbation producing the greatest perturbation growth over a specified interval of time (this perturbation is called the optimal perturbation). This initial perturbation can be obtained directly from analysis of the propagator, and it does not, in general, coincide with a particular mode of the system.

The optimal growth over time t is given by the norm of the propagator $||\mathbf{\Phi}(t)||$, which is defined to be the maximum measure of $\mathbf{x}(t)$ over all initial unit perturbations $\mathbf{x}(0)$ (the measures used at the initial and final times need not be the same). In the meteorological literature a systematic theory has been developed for identifying optimal perturbations in the L_2 norm [*Farrell*, 1988; *Farrell and Ioannou*, 1996a,b], but a comparable theory has not been advanced for chemical perturbations. As discussed above, the theory developed in meteorology can not be carried over directly to chemistry, because of the inherent distinction between the measure of perturbations in dynamical systems (the energy of perturbations), which is quadratic and associated with an L_2 norm, and the measure in chemical systems (the number of molecules), which is linear. (There is a notable exception: if we choose to measure the final perturbations only at a specific location, then the linear measure gives the same values as are obtained with the L_2 norm. In these cases the results available for analysis of nonnormal systems with L_2 norms [*Farrell and Ioannou*, 1996a] are directly applicable.) It turns out, however, that the linearity of the appropriate measure in chemical systems is, in fact, an advantage that leads to a particularly efficient method for obtaining optimal initial conditions. In the sequel the appropriate measure of chemical perturbations is introduced, and the theory for obtaining optimal perturbations to chemical systems is developed.

In order to streamline the development we will limit the analysis to time-independent systems, for which the propagator is the matrix exponential $\mathbf{\Phi}(\mathbf{t}) = e^{\mathbf{A}t}$. The analysis can be carried over, with minor modifications, to time-dependent systems.(The system propagator for a time-dependent system can be obtained by forward integration of the identity matrix. We note that eigenanalysis is not defined for time-dependent systems, so that methods of analysis based on eigenmodes can not be extended to apply to time-dependent systems.)

A simple atmospheric chemical system consisting of the advection, diffusion, loss, and interaction of CH_3Br and Br_y in a one-dimensional atmospheric column is taken as an example (the example is based on the B1 model of *Prather* [1977]). We apply to this system the methods of nonnormal analysis and obtain general properties of the solution including the response of the system to forcing applied impulsively, continuously, and stochastically distributed in space and time: the perturbations most effective in disturbing the chemical equilibrium and appropriate measures of timescale for this system. Before presenting the theory we describe the example chemical system.

2. CH_3Br/Br_y Perturbation Dynamics

The mixing ratios of a species, denoted χ_i, satisfy the advection equation [*Brasseur and Solomon*, 1986; *Andrews et al.*, 1987]:

$$\frac{\partial \chi_i}{\partial t} + \vec{\mathbf{u}} \cdot \nabla \chi_i = S \,, \tag{3}$$

where $\vec{\mathbf{u}}$ is the velocity field and S denotes the net sources. Consider a model of the ascending branch of the Hadley circulation at the equator. Zonal homogeneity and symmetry about the equator imply that the horizontal derivatives vanish. The advection equation that governs evolution of mixing ratios of the chemical constituents with height and time reduces under these assumptions to

$$\frac{\partial \chi_i}{\partial t} + w(z)\, \frac{\partial \chi_i}{\partial z} = S \,. \tag{4}$$

Using (4), evolution of number densities in the coupled CH_3Br/Br_y system satisfies the following equations:

$$\begin{aligned} \frac{dn_{CH_3Br}}{dt} &= -\, w(z)N(z)\, \frac{d}{dz}\left[n_{CH_3Br}/N(z)\right] + \\ &+ \frac{d}{dz}\left\{K(z)N(z)\, \frac{d}{dz}\left[n_{CH_3Br}/N(z)\right]\right\} - \\ &-\, m_1(z)\, n_{CH_3Br} + P_{CH_3Br}, \end{aligned} \tag{5a}$$

$$\begin{aligned} \frac{dn_{Br_y}}{dt} &= -\, w(z)N(z)\, \frac{d}{dz}\left[n_{Br_y}/N(z)\right] + \\ &+ m_1(z)\, n_{CH_3Br} - m_2(z)\, n_{Br_y} + \\ &+ \frac{d}{dz}\left\{K(z)N(z)\, \frac{d}{dz}\left[n_{Br_y}/N(z)\right]\right\} . \end{aligned} \tag{5b}$$

Here z is the height, n_i is the number density (m^{-3}) of species i, and $N(z)$ is the atmospheric number density (m^{-3}). The source of CH_3Br is P_{CH_3Br} ($m^{-3}\ s^{-1}$), which may vary both in space and in time. The chemistry is as in the *Prather* [1997] B1 model. The CH_3Br is destroyed at a rate $m_1(z)$ by reacting with OH in the troposphere ($z < 10$ km) and by photolysis in the stratosphere. The bromine species, collectively denoted by Br_y, are formed from the destruction of CH_3Br and rained out at a rate $m_2(z)$ in the troposphere. The vertical distribution of $m_1(z)$ and $m_2(z)$ is shown in Figure 1a. The eddy diffusion coefficient is denoted by K. The first term on the right hand side of (5a) and (5b) represents vertical advection by the mass-conserving velocity field. The distributions with height of the vertical velocity $w(z)$ and of the diffusion coefficient $K(z)$ are shown in Figure 1b.

The continuity equations for the number density tendencies (equations (5a) and (5b)) are collocated at equally spaced levels in z, and the spatial differential operators are approximated with second-order difference operators. The number of collocation points N is chosen such that the solutions are converged to the continuous system (typically 60 levels are adequate).

Equations (5a) and (5b) are written in matrix form:

$$\frac{d\mathbf{n}}{dt} = \mathbf{A}\mathbf{n} + \mathbf{p}, \tag{6}$$

where $\mathbf{n}$ is the column vector of the number densities in which the first N elements are the number densities of CH_3Br in sequence from the ground to the top of the

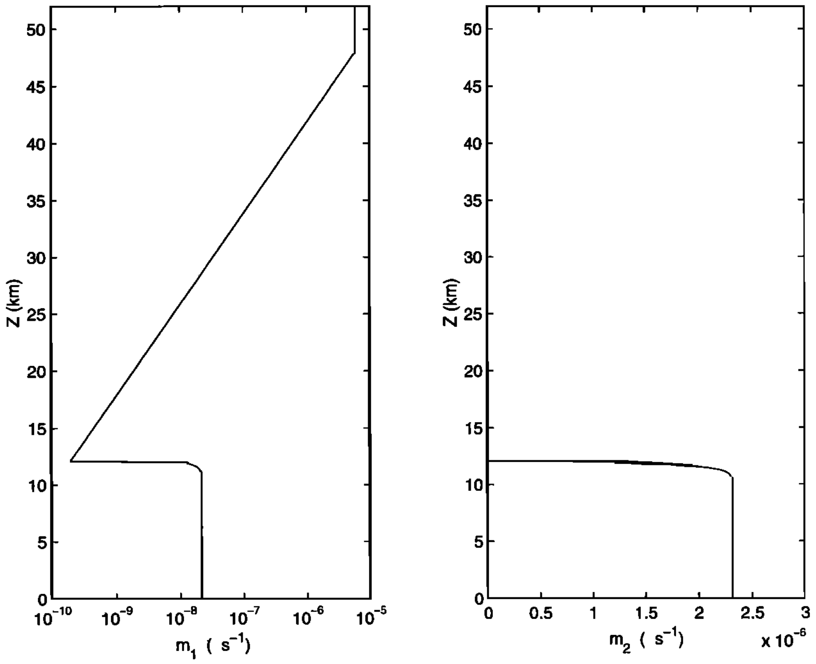


Figure 1a. (left) Chemical loss rate of CH_3Br as a function of height. The loss rate is constant in the troposphere at 2.166×10^{-8} s^{-1}; above 10 km it increases with height as $6 \times 10^{-6}\ p^{-2}$, where $p = 1000 \times 10^{-z/16}$ mbars is the pressure at height z (kilometers), and it is constant above $p = 1$ mbar. (right) Chemical loss rate of Br_y: 2.315×10^{-6} below 10 km owing to rain out, and zero elsewhere. Values are as in the work of *Prather* [1997].

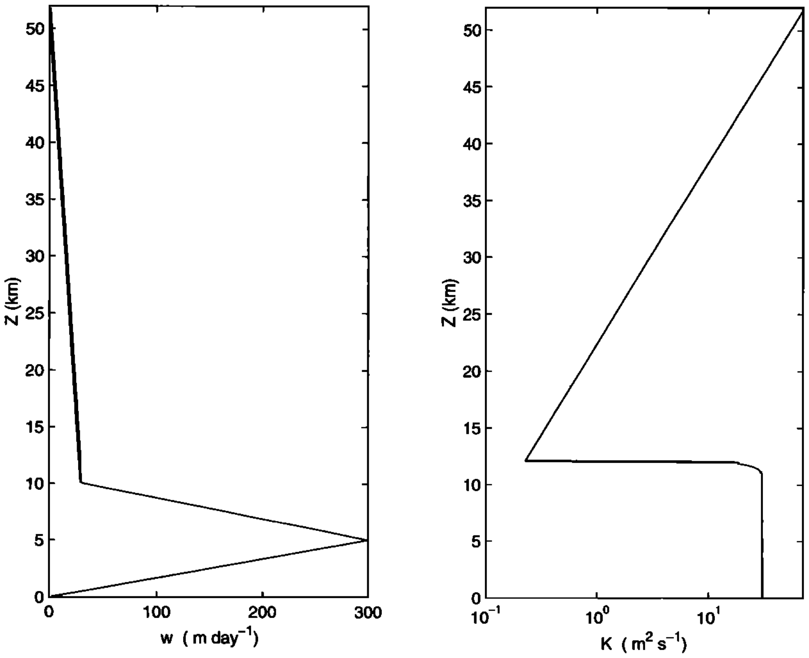


Figure 1b. (left) Vertical velocity (m d^{-1}) as a function of height (km). For $z \leq 5$, $w = 60z$, for $5 < z < 10$, $w = 300 - 54(z - 5)$, and for $z \geq 10$, $w = 30\,[1 - (z - 10)/42]$. (right) Diffusion coefficient as a function of height: $K = 30$ m^2 s^{-1} for $z < 12$, $K = 0.3$ m^2 s^{-1} at $z = 14$ km, and K increases as p_{14}/p above 14 km, where p is the atmospheric pressure and p_{14} is its value at 14 km (as in the work of *Prather* [1997]).

stratosphere (taken at 52 km), and the last N elements are the corresponding Br_y number densities. The source vector $\mathbf{p}$ is $[P, 0]^T$, as no source of Br_y is included. The matrix $\mathbf{A}$ has the form

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} , \quad (7)$$

where all the submatrices are $N \times N$; $\mathbf{A}_{11}$ is the matrix of the discretized differential operator

$$\mathbf{A}_{11} = -w(z)N(z)\frac{d}{dz}\frac{1}{N(z)} - m_1(z) + + \frac{d}{dz}\left(K(z)N(z)\frac{d}{dz}\frac{1}{N(z)}\right) , \quad (8)$$

operating on $\mathbf{n}_{CH_3Br}$; $\mathbf{A}_{12}$ is zero; $\mathbf{A}_{21}$ is the diagonal matrix with the values of $m_1(z)$; and finally $\mathbf{A}_{22}$ is

$$\mathbf{A}_{22} = -w(z)N(z)\frac{d}{dz}\frac{1}{N(z)} - m_2(z) + +\frac{d}{dz}\left(K(z)N(z)\frac{d}{dz}\frac{1}{N(z)}\right) , \quad (9)$$

operating on $\mathbf{n}_{Br_y}$.

The state vector of the number densities at time t is given by

$$\mathbf{n}(t) = e^{\mathbf{A}t}\, \mathbf{n}(0) + \int_0^t e^{\mathbf{A}(t-s)}\, \mathbf{p}(s)\, ds , \quad (10)$$

where $\mathbf{n}(0)$ is the initial distribution of the chemical species. When all the eigenvalues of $\mathbf{A}$ are stable, i.e., have negative real part, a well-defined steady state exists which is independent of the initial distribution $\mathbf{n}(0)$. For the simple case in which the source is independent of time this steady state is

$$\bar{\mathbf{n}} = -\mathbf{A}^{-1}\,\mathbf{p} , \quad (11)$$

which can be converted into mixing ratios by multiplying $\mathbf{n}$ by the $2N$ diagonal matrix $\mathbf{M}$ with the following elements:

$$M_{ij} = \begin{cases} \delta_{ij}/N(z_i) & i,j = 1,\cdots,N \\ \delta_{ij}/N(z_{i-n}) & i,j = N+1,\cdots,2N . \end{cases} \quad (12)$$

The time development of perturbations about this mean state is determined by the free solution in (10). Because $\mathbf{A}$ is time independent, the propagator (see equation (2)) has the form of the matrix exponential, i.e., $\mathbf{\Phi}(t) = e^{\mathbf{A}t}$.

3. Measures of Chemical Abundance

In order to proceed with the study of perturbation growth, it is necessary to adopt a measure to quantify species abundances in the initial state and a measure to quantify species abundances at a later time. These two measures need not be the same.

In many dynamical applications the appropriate measure is the euclidean or L_2 norm

$$||\mathbf{n}||_2 = \left\{\sum_{i=1}^{N} \Delta z \left[\mathbf{n}^2_{CH_3Br}(z_i) + \mathbf{n}^2_{Br_y}(z_i)\right]\right\}^{1/2} , \quad (13)$$

which is a rms measure: for chemical systems it is the euclidean length of the vector of species abundances at grid levels. Stability theory for nonnormal dynamical systems in the L_2 norm is well developed [cf. *Farrell and Ioannou*, 1996a,b].

However, the L_2 norm is not well suited for addressing questions arising in chemistry. For chemistry problems, physically appropriate perturbation measures are usually linear in the number density or mixing ratios of the constituents. One such measure is what we will call the integral measure:

$$\Delta z \left|\sum_{i=1}^{N} w_{1i}\, \mathbf{n}_{CH_3Br}(z_i) + w_{2i}\, \mathbf{n}_{Br_y}(z_i)\right| , \quad (14)$$

with $\mathbf{w}$ as a weighting factor. This measure is a weighted column abundance of the species. The weights can be chosen so that the abundances of chosen species at chosen locations contribute to the measure. An example application of (14) is to the measurement of perturbations in Br_y mixing ratio at specified locations in the stratosphere which may be of interest for assessing ozone depletion potential. For this example, w_{1i} are taken as zero and w_{2i} are taken as nonzero only at the levels of interest for the assessment. Note that the integral measure is not a norm, because there exist nonzero perturbations for which the integral measure is zero. (A vector norm is a function f from R^n to R that satisfies the following requirements: $f(x) \geq 0$ and $f(x) = 0$ if and only if $x = 0$, $f(x+y) \leq f(x) + f(y)$, and $f(\alpha x) = |\alpha| f(x)$ for all vectors x, y, and all scalars α.)

It is convenient to introduce a weighting matrix $\mathbf{W}$ to define the most general linear integral measure of chemical perturbations:

$$\left[\mathbf{n}\right]_{\mathbf{W}} \equiv \Delta z \left|\sum_{i=1}^{N_s} \sum_{j=1}^{N_s} W_{ij}\, \mathbf{n}_i\right| , \quad (15)$$

where $\mathbf{W}$ is the species measurement weighting matrix and N_s is the total dimension of the system (in the example above $N_s = 2N$ and $\mathbf{W}$ is diagonal with elements along the diagonal equal to the $\mathbf{w}$ values in (14)). When all species abundances contribute equally to the measure, $\mathbf{W} = \mathbf{I}$ (where $\mathbf{I}$ is the identity) and the integral measure will be denoted $[\mathbf{n}]$, without reference to the weighting matrix. (Note that the identity integral measure of the vector $\mathbf{Wn}$ is equal to the $\mathbf{W}$ integral measure of $\mathbf{n}$, i.e., $[\mathbf{Wn}] = [\mathbf{n}]_{\mathbf{W}}$. As a result, only the identity integral measure is needed: when a state $\mathbf{n}$

is to be measured with weighting matrix $\mathbf{W}$, it suffices to calculate the identity integral measure of $\mathbf{Wn}$.)

The linear measure with weighting matrix $\mathbf{W}$ is used to quantify abundances at a time t after the introduction of the initial perturbation. While the final state is measured with the integral measure, which need not be a norm, it is not sensible to measure the initial state in the same manner. The reason is that the integral measure can assign zero measure to nonzero perturbations, because initial perturbations of both signs can be introduced which have zero integral measure. It is necessary to measure the initial perturbations with a norm, i.e., with a measure that will give positive magnitude to all nonzero perturbations. We will employ a norm consisting of a weighted sum of the absolute values of mixing ratios (or number densities depending on context) defined as

$$||\mathbf{n}||_1^{\mathbf{W}} = \Delta z \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} \left| W_{ij}\, \mathbf{n}_i \right| , \qquad (16)$$

which is the L_1 norm of the weighted vector $\mathbf{Wn}$. (The L_1 norm of a vector is equal to the sum of the absolute value of its entries, i.e., $||\mathbf{n}||_1 = \sum_i |\mathbf{n}_i|$.)

It is common in chemistry problems for the solution fields to be of one sign for initial perturbations of one sign (which requires that the propagator matrix $\mathbf{\Phi}(t)$ have all its elements of one sign). In these cases the integral measure given in (15) is equivalent to the L_1 norm, as negative perturbations can not arise if not introduced initially, so that the absolute value is redundant in (15). For example, the CH_3Br/Br_y system has a propagator with all positive elements for which the L_1 norm is appropriate.

4. Determining Optimal Growth

Having fixed the measures with which to quantify chemical perturbations, we consider the initial perturbation that produces the greatest net change at time t (the proofs and details of the methods can be found in appendix A). The ratio of the abundances at time t to the initial abundance will be called the growth at time t. The maximization problem we consider is to determine the initial state $\mathbf{n}(0)$ of unit norm, i.e., $||\mathbf{n}(0)||_1 = 1$, producing the state of maximum norm at time t, i.e.,

$$\begin{aligned} \max_{||\mathbf{n}(0)||_1=1} \Big[\mathbf{n}(t)\Big]_{\mathbf{W}^f} &= \max_{||\mathbf{n}(0)||_1=1} \Big[\mathbf{W}^f \mathbf{\Phi}(t)\mathbf{n}(0)\Big] \\ &\equiv ||\mathbf{W}^f \mathbf{\Phi}(t)|| \ , \qquad (17) \end{aligned}$$

where $\mathbf{W}^f$ is a measuring weighting matrix of the final state (FSW). The optimal growth in this case is denoted $||\mathbf{W}^f \mathbf{\Phi}(t)||$ in order to indicate that it depends on the propagator at the optimization time t and on the final weighting matrix $\mathbf{W}^f$.

We may want to treat a more general optimization problem in which there is weighting imposed also on the initial states in order, for example, to emphasize certain chemical constituents more than others, or to emphasize initial perturbations in selected regions. This can be done by formulating the maximization problem of determining the initial perturbation $\mathbf{n}(0)$ of unit norm that maximizes $[\mathbf{W}^f \mathbf{\Phi}(\mathbf{t})\mathbf{W}^i \mathbf{n}(0)]$, where $\mathbf{W}^i$ is a weighting matrix of the initial state (ISW). By this means we consider all initial perturbations and transfer the biases in their relative importance to the weighting matrix $\mathbf{W}^i$. For example, in the CH_3Br/Br_y system we can allow initial perturbations only in CH_3Br, in which case $\mathbf{W}^i$ is the diagonal matrix $w_{ai}\delta_{ij}$ ($a = 1,2$), with $w_{1i} = 1$, $i = 1, \cdots N$, and $w_{2i} = 0$ for all i.

The initial perturbation producing the maximum perturbation at time t, called the optimization time, is called the optimal perturbation, and the associated maximum response resulting at time t is called the evolved optimal, and its measure is called the optimal growth. The optimal growth at time t is defined as

$$||\mathbf{W}^f \mathbf{\Phi}(t)\mathbf{W}^i|| = \max_{||\mathbf{n}(0)||_1=1} \Big[\mathbf{W}^f \mathbf{\Phi}(t)\mathbf{W}^i \mathbf{n}(0)\Big], \qquad (18)$$

where the notation $\max_{||\mathbf{n}(0)||_1=1}$ denotes maximization over initial perturbations $\mathbf{n}(0)$ of unit norm and the optimal growth is denoted as $||\mathbf{W}^f\ \mathbf{\Phi}(t)\ \mathbf{W}^i||$ to indicate that it depends on the propagator at the optimization time t and on the final and initial state weighting matrices $\mathbf{W}^f$ and $\mathbf{W}^i$.

Remarkably, the optimal perturbation so defined consists of excitation of a single species at the level corresponding to the column of greatest absolute sum of $\mathbf{W}^f\ \mathbf{\Phi}(t)\ \mathbf{W}^i$, and the optimal growth is this sum (appendix A). In the continuous limit the optimal perturbation approaches a delta function of a single species at a specific level. This delta function subsequently evolves with time as the Green's function of the operator. A similar result is obtained if the final state is measured in the L_1 norm, except in this case the optimal perturbation consists of excitation of a single species at the level corresponding to the column of $\mathbf{W}^f\ \mathbf{\Phi}(t)\ \mathbf{W}^i$ with the greatest sum of the absolute values of its elements, and the optimal growth is this sum.

The same magnitude of perturbation growth is obtained for the positive and the negative delta functions: in the first case we have identified the perturbation with growth of greatest positive measure, while in the second case we have identified the perturbation with growth of greatest negative measure. In general, the initial perturbation producing the final perturbation of minimum magnitude in the integral measure is not a spatial delta function of a single species. However, the minimum growth in magnitude in the L_1 norm is a delta function of a single species at the single location corresponding to the column of $\mathbf{W}^f\ \mathbf{\Phi}(t)\ \mathbf{W}^i$ with minimum sum of the absolute value of its elements. We can exploit this simple result when minimum impact is sought in the integral measure even in cases when positivity of the unrestricted propagator is not assured, because it is often possible to choose an appropriate FSW to ensure positivity and allow the minimum to be identified with a delta function initial perturbation as the integral measure then becomes the L_1 norm.

It is a remarkable property that if the propagator is known, then we can determine the optimal perturbation in the integral measure by determining the column of the propagator with the maximum absolute sum. While for small-dimensional examples, obtaining the propagator with matrix manipulations is certainly possible, this is impossible in large global chemical models. The propagator could be obtained in that case by integrating forward the identity matrix, which requires a number of integrations equal to the dimensionality of the system. However, this is a numerically expensive procedure. In meteorology where L_2 optimal perturbations are obtained routinely in the course of operational forecasting, this problem is solved by singular value decomposition using iterative techniques based on forward time integration of the model, followed by backward integration of the adjoint model until a subset of optimals and evolved optimals has been found. Because of the linearity of the measure in chemical models, a great simplification occurs, and chemical optimal perturbations can be determined by a single integration of the transposed dynamical system.

This result follows immediately from the observation that the transposed dynamical system, with dynamical operator $\mathbf{A}^T$, has as its propagator $\mathbf{\Phi}(t)^T$, as $(e^{\mathbf{A}t})^T = e^{\mathbf{A}^Tt}$. Consider the initial condition $\mathbf{n}(0) = [1,1,\cdots,1]^T$. This initial condition will evolve to

$$\tilde{\mathbf{n}}(t) = \mathbf{\Phi}(t)^T\mathbf{n}(0) , \qquad (19)$$

which is a vector with entries the sums of the columns of $\mathbf{\Phi}(t)$. The entry of $\tilde{\mathbf{n}}(t)$ with the maximum value determines the optimal growth at time t. If this maximum value occurs at the kth entry of $\tilde{\mathbf{n}}(t)$, then the optimal perturbation is the unit vector $\mathbf{n}_{\text{opt}}$, with entries $(\mathbf{n}_{\text{opt}})_\imath = \delta_{\imath k}$. The evolved optimal is then obtained by forward integration to time t of the system with initial condition $\mathbf{n}_{\text{opt}}$. Therefore, without obtaining the propagator we can determine the optimal perturbation with a single adjoint integration and the evolved optimal with one more integration of the system.

If, now, we want to determine the perturbation that leads to the optimal growth $||\mathbf{W}^f \; \mathbf{\Phi}(t) \; \mathbf{W}^\imath||$, we consider the initial condition

$$\mathbf{n}(0) = (\mathbf{W}^f)^T \; [1,1,\cdots,1]^T \; , \qquad (20)$$

which we integrate forward by the transpose system $\mathbf{A}^T$ to obtain

$$\tilde{\mathbf{n}}(t) = \mathbf{\Phi}(t)^T\mathbf{n}(0) \; . \qquad (21)$$

The maximum entry of $(\mathbf{W}^\imath)^T\tilde{\mathbf{n}}(t)$ is the optimal growth. If the maximum occurs at the kth entry of $(\mathbf{W}^\imath)^T\tilde{\mathbf{n}}(t)$, then the optimal perturbation is the unit vector $\mathbf{n}_{\text{opt}}$, with entries $(\mathbf{n}_{\text{opt}})_\imath = \delta_{\imath k}$. The evolved optimal is then obtained by forward integration to time t of the system with initial condition $\mathbf{n}_{\text{opt}}$.

In the asymptotic limit $t \to \infty$, all initial perturbations assume the structure of the least damped mode and decay at a rate given by the real part of the least damped mode eigenvalue. This suggests the problem of determining the initial perturbation which optimally excites the least damped mode and consequently produces the greatest disturbance to the system as $t \to \infty$. In the integral measure it turns out that the initial perturbation that optimally excites the least damped mode is a delta function of a single species located at the level corresponding to the maximum amplitude not of the mode itself but of its quite distinct bi-orthogonal (see section 1). The proof of this theorem is given in appendix B. The structure of the bi-orthogonal differs from that of the mode in nonnormal systems. The bi-orthogonal can be obtained either by inverting the matrix of the eigenvectors or by eigenanalysis of the transpose operator $\mathbf{A}^T$.

Often it is required to maximize the integrated concentration of a species at a chosen location after a pulsed emission. For example, the ozone depletion potential is given by the time integral of Br_y concentration in the stratosphere [*Prather*, 1996, 1997]. It is thus required to obtain

$$||\mathbf{W}^f \left(\int_0^\infty e^{\mathbf{A}t} dt \right) \mathbf{W}^i|| = || - \mathbf{W}^f \mathbf{A}^{-1} \mathbf{W}^i||, \qquad (22)$$

for some initial weighting $\mathbf{W}^i$ and final weighting $\mathbf{W}^f$. The solution to this problem is to locate the pulse at a level corresponding to the column of $\mathbf{W}^f \; \mathbf{A}^{-1} \; \mathbf{W}^\imath$ with the maximum absolute sum.

5. Example

As an example, consider the CH_3Br/Br_y system. The eigenmodes of this system can be separated into two classes: modes that involve only Br_y perturbations (the Br modes) and those that involve both CH_3Br and Br_y perturbations (the CH_3Br modes) [*Prather*, 1997]. The least damped CH_3Br mode with only diffusion (no advection) is shown as curve 1 in Figure 2; it has a decay time of 2.1 years. Its bi-orthogonal (for calculation of the bi-orthogonal see section 1) consists of only CH_3Br perturbation and is also shown in Figure 2 (curve 2). Optimal excitation in the integral measure of this least damped CH_3Br mode is obtained by placing a CH_3Br perturbation in number density at 19.4 km, where the bi-orthogonal reaches its maximum, which results in approximately a twofold increase in amplitude of the mode over introduction of the mode itself as an initial perturbation. Advection reduces the decay times and shifts the maximum amplitude of the mode to higher altitudes while its bi-orthogonal is shifted toward the ground as can be shown in Figure 3 for the case of maximum advection $w = 300$ m d^{-1} in the midtroposphere and $w = 30$ m d^{-1} in the stratosphere (see Figure 1b).

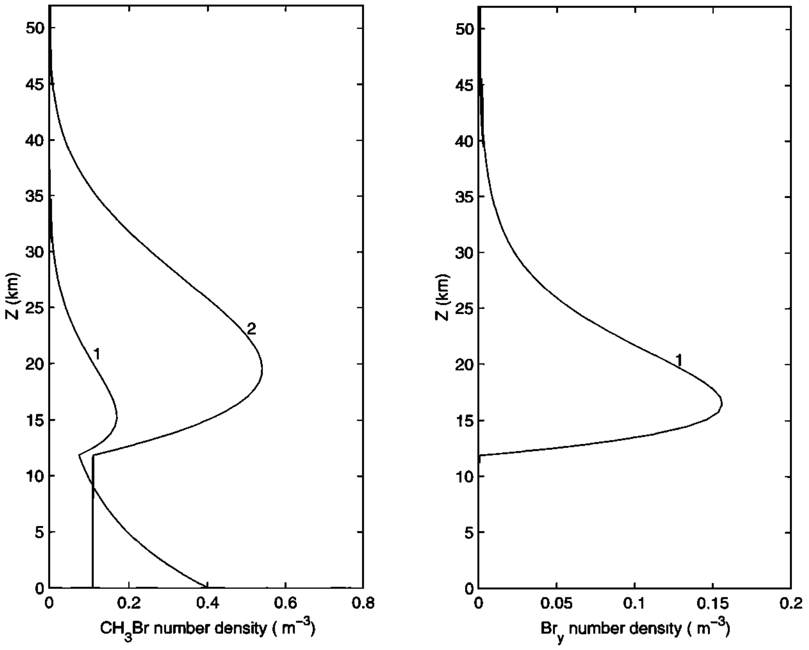


Figure 2. (left) CH_3Br number density. (right) Br_y number density. Curve 1 shows the vertical structure of the least damped CH_3Br mode for zero advection (decay time 2.1 years). Curve 2 shows the bi-orthogonal of this mode which comprises only a CH_3Br perturbation. Optimal excitation of the mode in the integral measure is achieved by placing unit CH_3Br perturbation at 19.4 km where the bi-orthogonal has its maximum. The modes have been normalized to unity in L_2.

For this case the least damped CH_3Br mode (curve 1) has a decay time of 1.48 year. Its bi-orthogonal (curve 2) is again a perturbation in CH_3Br alone. The optimal excitation is impulsive introduction of CH_3Br at the ground, which leads to a factor 1.6 increase in the amplitude of the mode compared with the amplitude that would have been obtained if the mode itself were introduced initially. Examination of the structure of the bi-orthogonals in Figures 2 and 3 reveals that the least damped mode can be effectively excited by CH_3Br perturbations in the stratosphere in the example with diffusion only (Figure 2) while inclusion of advection in the dynamics renders CH_3Br perturbations in the stratosphere totally ineffective in exciting the least damped mode (Figure 3).

As an example of optimal excitation, we consider the problem of finding the initial perturbations with the greatest ozone depletion potential which requires a FSW that identifies the mixing ratio concentrations of Br_y in the lower stratosphere. For that purpose a suitable final state weighting is $\mathbf{W}^f = \mathbf{P}\ \mathbf{M}$, where $\mathbf{M}$ is the matrix (equation (12)) that converts number densities into mixing ratios and $\mathbf{P}$ projects the final state to the Br_y species at 20 km; that is, $P_{kl} = \delta_{l\alpha}\delta_{kl}$ (no summation), where α is the index of Br_y at the level of 20 km. We allow initial number density perturbations only in CH_3Br, so the initial state weighting is $W^i_{kl} = \delta_{kl}$ for the CH_3Br entries $(k, l \leq N)$ and zero otherwise. We normalize the initial CH_3Br to correspond to the number density that would result in a mixing ratio of 1 parts per trillion (ppt) over the first 2 km above the ground. The optimal growth at time t is given by the integral measure of the propagator $||\mathbf{W}^f e^{\mathbf{A}t}\mathbf{W}^i||$ (see equation (18)), which because of positivity of the propagator is the same as its L_1 norm which equals the greatest absolute column sum of the matrix $\mathbf{W}^f e^{\mathbf{A}t}\mathbf{W}^i$.

The optimal mixing ratio of Br_y at 20 km and the location of the initial CH_3Br perturbation as a function of time are shown in Figure 4 for the case in which both advection and diffusion are present. The evolution of two optimal initial conditions is also shown: the optimal for 6 years and the optimal for 2 months. For large times (in this model for $t > 5.5$ years), optimal mixing ratios of Br_y are obtained by placing the CH_3Br perturbation at the ground. For intermediate times it is best to place the CH_3Br perturbation at the top of the stratosphere, while for the shortest times the optimal is near the target level. The level of injection that would result in the least CH_3Br response is also shown in figure 4. The maximum possible growth is called the global optimal, and it occurs at a time which will be referred to as a global optimal time. For the chosen parameters this is 1 year. Consequently, at least for a year following the introduction of a perturbation, decay time estimates based on the inverse of the decay rate of the least damped mode are not valid. In fact, for the chosen parameters the least damped mode decay rate (decay time of 1.48 years) is not obtained until 4 years after an impulsive excitation.

The global optimal time decreases with increasing ad-

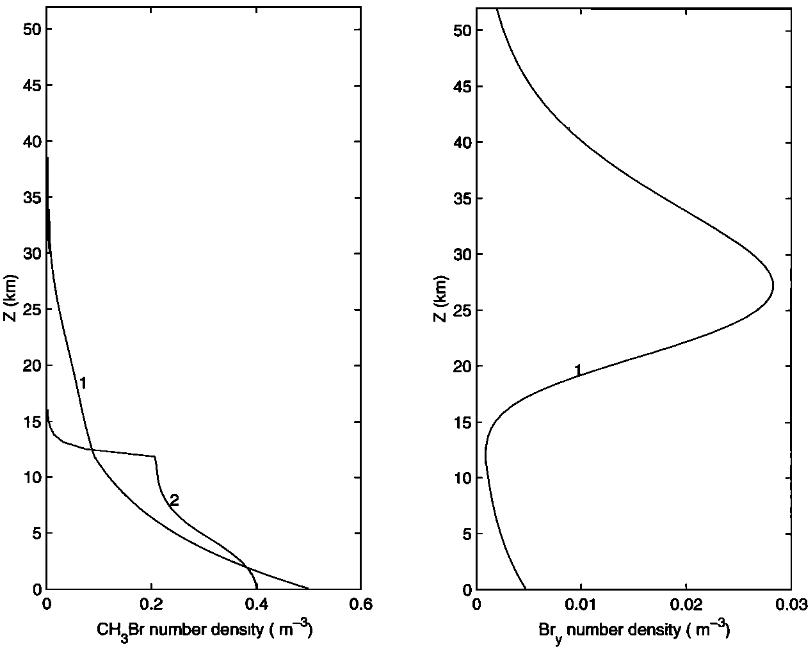


Figure 3. (left) CH_3Br number density. (right) Br_y number density. Curve 1 shows the vertical structure of the least damped CH_3Br mode with advection. The decay time of the mode is 1.48 years. Curve 2 shows the bi-orthogonal of this mode which comprises only a CH_3Br perturbation. Optimal excitation of the mode in the integral measure is achieved by placing unit CH_3Br perturbation at the ground where the bi-orthogonal has its maximum. The modes have been normalized to unity in L_2.

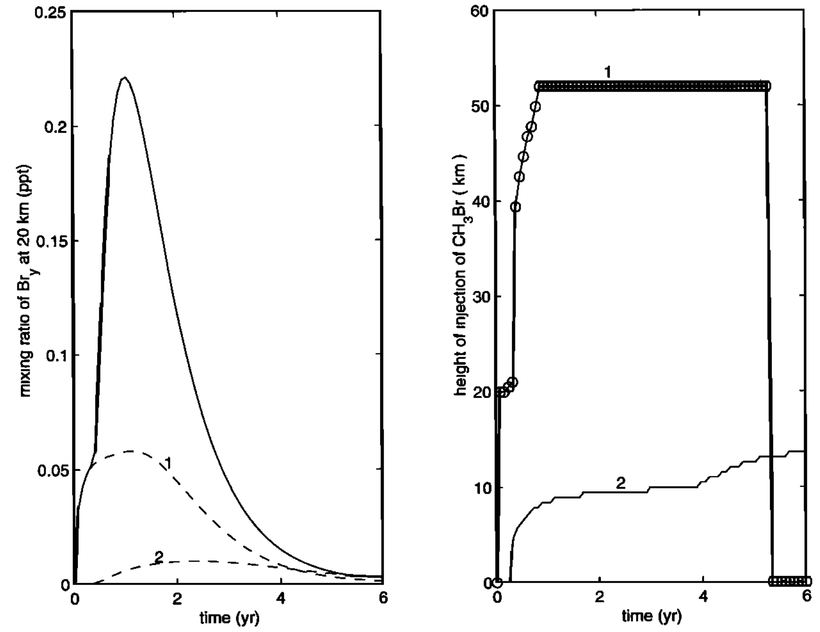


Figure 4. (left) Optimal growth of Br_y mixing ratio (parts per trillion (ppt)) at 20 km as a function of optimization time. Optimization time is defined as the time selected to calculate the maximum possible growth (the optimal growth) in the system as quantified by the selected measure of chemical perturbations, which is here the mixing ratio of Br_y at 20 km. Only CH_3Br initial perturbations are considered. The amount released at the ground over a grid interval Δz is equivalent to 1 ppt over the first 2 km (4.8×10^{16} m^{-2}). Both advection and diffusion are included. Dashed curve 1 shows the time evolution of the 2 month optimal perturbation. Dashed curve 2 shows the time evolution of the 6 year optimal perturbation, which is initially at the ground. (right) Height (kilometers) of the optimal CH_3Br perturbation is shown as a function of optimization time (curve 1). Height (kilometers) of the initial perturbation in CH_3Br that leads to the least response in Br_y mixing ratio at 20 km is shown as a function of optimization time (curve 2).

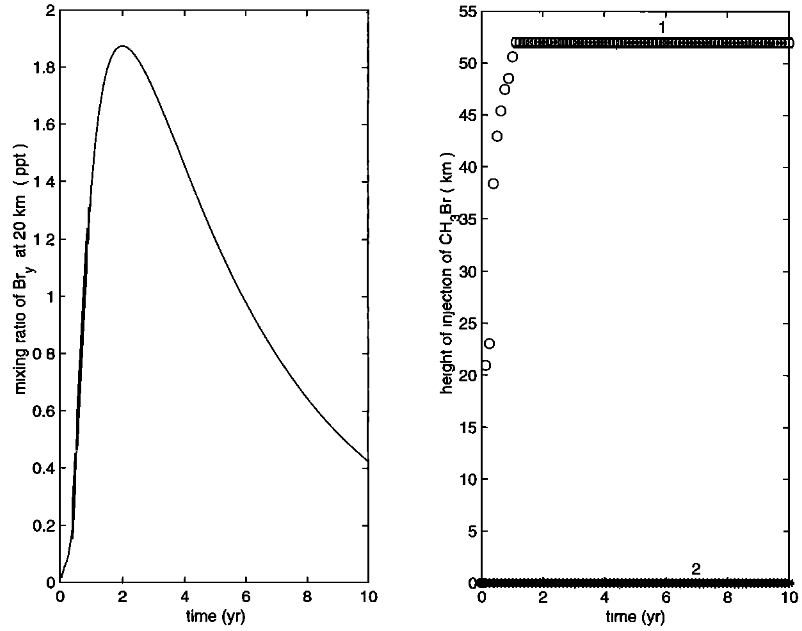


Figure 5. (left) Optimal growth of Br_y mixing ratio (ppt) at 20 km as a function of optimization time. Optimization time is defined as the time selected to calculate the maximum possible growth (the optimal growth) in the system as quantified by the selected measure of chemical perturbations which is here the mixing ratio of Br_y at 20 km. Only CH3Br initial perturbations are considered. The amount released at the ground over a grid interval Δz is equivalent to 1 ppt over the first 2 km (4.8×10^{16} m^{-2}). Only diffusion is included (no advection). (right) Height (kilometers) of the optimal CH3Br perturbation is shown as a function of optimization time (curve 1). Height (kilometers) of the initial perturbation in CH3Br that leads to the least response in Br_y mixing ratio at 20 km is shown as a function of optimization time (curve 2).

vection, while the qualitative features of the optimal excitation remain. The optimal growth and levels of optimal excitation for different optimizing times with no advection are shown in Figure 5. The optimal level for CH3Br perturbations is always aloft, while perturbations at the ground produce the minimal response. The global optimal time is about 2 years and the asymptotic decay time of 2.1 years is obtained after 5 years. Again, estimates of species lifetime based on the asymptotic decay of the least damped mode of the system are valid only after the transient buildup phase has ended, which in our example requires 4 years.

6. Lifetimes

Traditional methods for analyzing response of chemical equilibria to perturbation proceed from the concept of a representative lifetime. If a chemical equilibrium concentration $n(t)$ were governed by a scalar differential equation of the form

$$\frac{dn}{dt} = -\alpha n + p , \tag{23}$$

then an initial perturbation $n(0)$ would decay back toward the equilibrium concentration p/α as

$$n(t) = e^{-\alpha t}\, n(0) + \frac{p}{\alpha} , \tag{24}$$

and a decay timescale $\tau_d = 1/\alpha$ characterizing the return of the system to equilibrium is well founded. Also, the reservoir turnover time given by the steady state abundance $n(\infty) = p/\alpha$ divided by the forcing p, $\tau_r = n(\infty)/p = 1/\alpha$, is equal to τ_d and is also well founded. However, if the system is multidimensional, then there are as many such timescales as the dimension of the system, because each mode has an associated timescale given by the reciprocal of its damping rate.

To further examine the concept of a representative lifetime, we first obtain a reservoir turnover time for our nonnormal system by solving the forced equation

$$\frac{d\mathbf{n}}{dt} = \mathbf{A}\mathbf{n} + \mathbf{p}(t) , \tag{25}$$

for a time-varying forcing $\mathbf{p}(t)$ with mean $\bar{\mathbf{p}}$. The mean state forced response (appendix D) is

$$\bar{\mathbf{n}} = -\,\mathbf{A}^{-1}\,\bar{\mathbf{p}} . \tag{26}$$

If the forcing is confined to CH3Br at the ground level and the amplitude of the forcing is chosen to produce 10 ppt at the ground at equilibrium, the resulting profile is as shown in Figure 6. The reservoir timescale is found by taking the ratio of the total abundance (m^{-3}) to the input rate (m^{-3} s^{-1}):

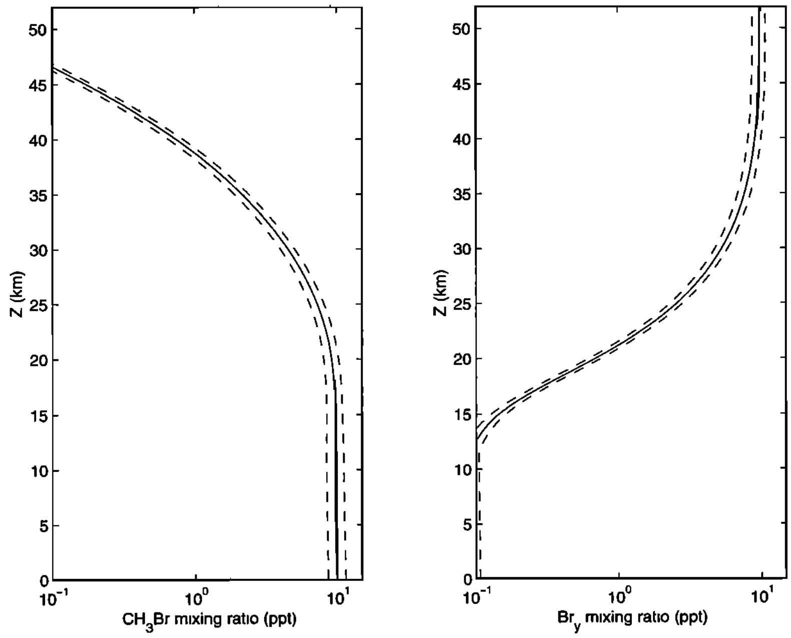


Figure 6. Mixing ratio maintained by a fluctuating source of CH_3Br at the ground. The source has a mean of 0.09 ppt d^{-1} and a variance of 1 ppt^2 $(12\ min)^{-2}$. Vertical advection is included. (left) Vertical distribution of mean CH_3Br. (right) Vertical distribution of Br_y. The dashed curves signify one standard deviation from the mean.

$$\tau_r = \frac{||\bar{\mathbf{n}}||_1}{||\bar{\mathbf{p}}||_1} . \tag{27}$$

This timescale is sensitive to the location of the forcing and to the nonnormality of $\mathbf{A}$ (for an explanation, see appendix D). The result of moving the forcing is shown in Figure 7. The reservoir timescale does not accurately reflect either the decay rate of transient perturbations at short time or the asymptotic decay rate at long times.

An alternative timescale is the relaxation timescale obtained after removal of the steady forcing maintaining the equilibrium state. This produces a continual variation of decay rates, implying decay times varying as

$$\tau(t) = \left[\frac{d||e^{\mathbf{A}t}\,(\mathbf{A}^{-1}\bar{\mathbf{p}})||_1}{dt} \right]^{-1} , \tag{28}$$

finally approaching the decay time of the least damped mode. This timescale depends on time elapsed since removal of the forcing, on the location of the forcing, and on the nonnormality of the evolution operator through the condition number of $\mathbf{A}^{-1}$. An example for forcing of CH_3Br at the ground is shown in Figure 8. Relaxation from the equilibrium state of a normal system proceeds with the instantaneous decay times monotonically increasing and asymptotically approaching the decay time of the least damped mode. Examination of Figure 9 reveals that this is not the case for the evolution of either CH_3Br or Br_y in the advecting case and is true only for relaxation of CH_3Br in the purely diffusive case. This reveals that the dynamics of number density evolution of both species in the presence of advection are dominated by nonnormality while the number density evolution of CH_3Br in the diffusive case is essentially that of a normal system. The degree of nonnormality depends on the variable we choose to study. As an example, the relaxation timescales for mixing ratios in the purely diffusive case are compared in Figure 9 to the relaxation times for number density. It is clear that the timescales depend on the choice of variable and also that the dynamics of mixing ratios have a greater degree of nonnormality than do the dynamics of number densities. The relaxation of CH_3Br is determined by a highly nonnormal operator when the variable is mixing ratio while when the variable is number density the governing operator is essentially normal. As a result, the decay times for mixing ratio are initially greatly increased compared with the decay times for number density. However, if one waits long enough for the asymptotic regime to be obtained, then the relaxation timescale becomes the decay time of the least damped mode which is sometimes thought to be the sole fundamental timescale of the system.

A third timescale is obtained from observing the size of fluctuations in the system. The fluctuation variance maintained by forcing CH_3Br stochastically at the ground is shown in Figure 6 (appendix C). In analogy with a scalar equation,

$$\frac{dn}{dt} = -\alpha n + \gamma \epsilon_t , \tag{29}$$

with ϵ_t as a white noise forcing with zero mean and unit variance. The maintained variance is

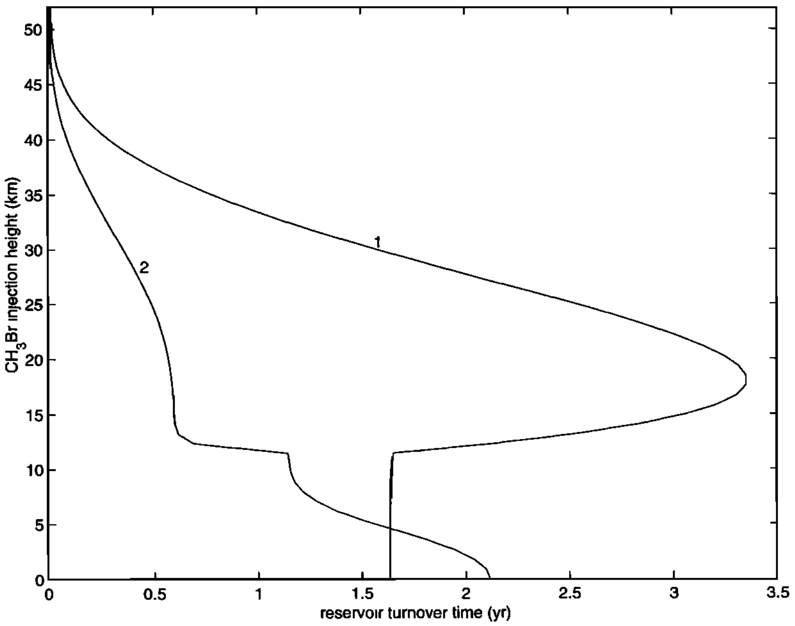


Figure 7. Reservoir turnover time as a function of the CH_3Br source height. Curve 1 shows turnover times with only diffusion included. Curve 2 shows turnover times with both advection and diffusion included.

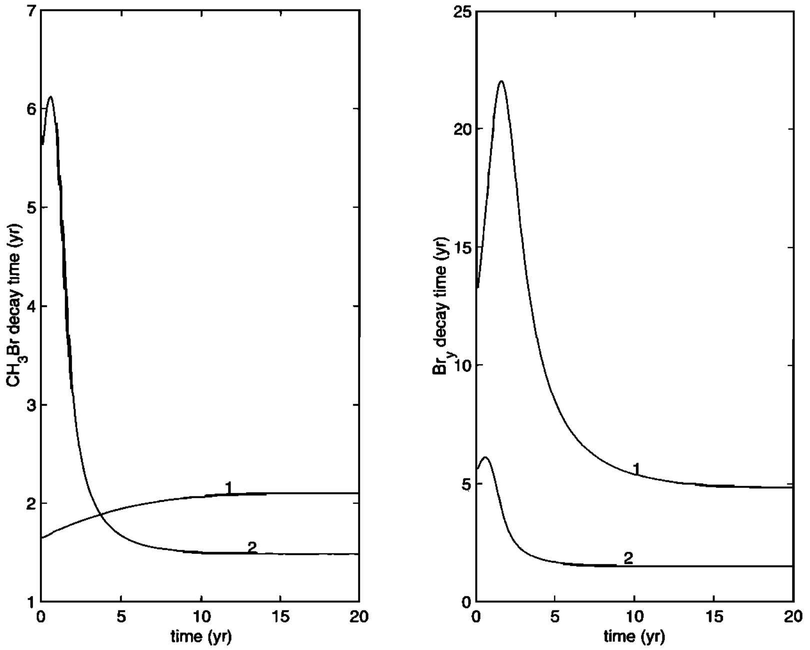


Figure 8. Timescales for relaxation from the equilibrium produced by a constant CH_3Br source at the ground. (left) Instantaneous decay rate of CH_3Br expressed as a timescale as a function of time from removal of the source. Curve 1 shows relaxation timescales with only diffusion included. Curve 2 shows relaxation timescales with both advection and diffusion included. (right) Instantaneous decay rate of Br_y expressed as a timescale as a function of time from removal of the source. Curve 1 shows instantaneous decay rates with only diffusion included. Curve 2 shows instantaneous decay rates with both advection and diffusion included. In all cases the decay time approaches that of the least damped mode for large t, but for an interval of time of approximately 5 years after removal of the source, the decay time differs substantially from that of the least damped mode.

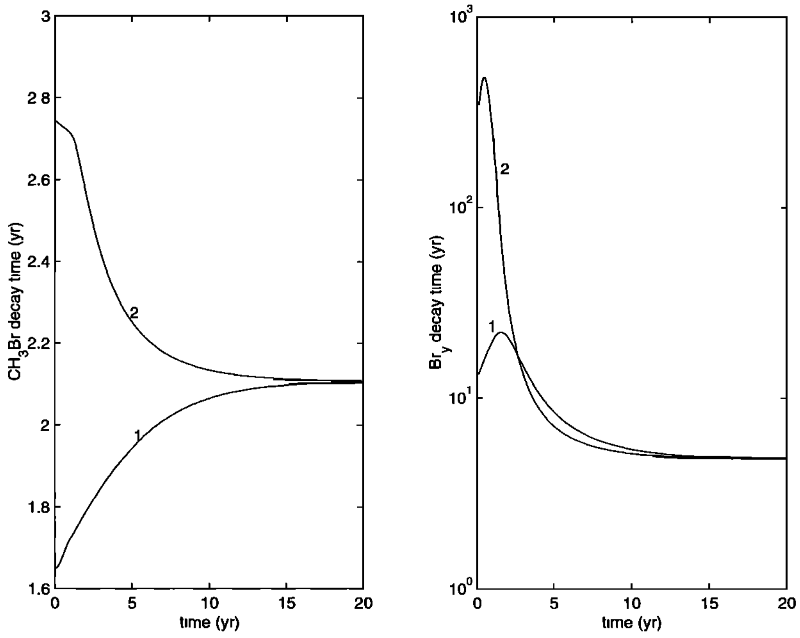


Figure 9. Relaxation timescales from an equilibrium state produced by a constant CH3Br source at the ground for the case of no advection. The influence of variable choice is revealed by the contrast between two cases. Curve 1 shows the instantaneous decay time of number density as a function of time from removal of the source. (left) CH3Br. (right) Br$_y$. Curve 2 shows the instantaneous decay time of mixing ratio as a function of time from removal of the source.

$$< n^2 > \; = \; \frac{\gamma^2}{2\alpha}, \tag{30}$$

from which the timescale

$$\tau_f \; = \; \frac{2 < n^2 >}{\gamma^2} \tag{31}$$

can be estimated. For our example the mean mixing ratio variance in the troposphere together with the forcing at the ground yield from (31) a timescale of 25 days. This estimated timescale is short because white noise forcing projects substantially on modes which are damped more rapidly than the modes primarily excited by the zero-frequency mean forcing. This frequency dependence of the timescale can be examined further by Fourier analysis of the perturbation version of (25):

$$\frac{d\mathbf{n}'}{dt} \; = \; \mathbf{A}\mathbf{n}' \; + \; \mathbf{p}'(t), \tag{32}$$

where $\mathbf{p}'(t) = \mathbf{p}(t) - \bar{\mathbf{p}}$; with transformed variable

$$\hat{\mathbf{n}}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{n}'(t) e^{-i\omega t} dt \, , \tag{33}$$

the response at frequency ω for the transformed forcings,

$$\hat{\mathbf{p}}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{p}'(t) e^{-i\omega t} dt \, , \tag{34}$$

weighted by $\mathbf{W}^i$ can be expressed as

$$\hat{\mathbf{n}}(\omega) \; = \; \mathbf{R}(\omega) \; \mathbf{W}^i \; \hat{\mathbf{p}}(\omega) \, , \tag{35}$$

in terms of the resolvent:

$$\mathbf{R} \; = \; (i\omega\mathbf{I} - \mathbf{A})^{-1} \, , \tag{36}$$

where $\mathbf{I}$ is the identity. In analogy with (21) we can define a damping time as a function of frequency to be the ratio of the total abundance (m^{-3}) to the input rate (m^{-3} s^{-1}). For sinusoidal forcing at frequency ω this timescale is

$$\tau_\omega \; = \; \frac{\| \; \hat{\mathbf{n}}(\omega) \; \|_1}{\| \; \mathbf{W}^i\hat{\mathbf{p}}(\omega) \; \|_1} \; . \tag{37}$$

At zero frequency this damping time is identical to the reservoir turnover time, but the damping time dramatically decreases with increasing frequency as shown in Figure 10. Consequently, timescales based on reservoir turnover time overestimate the persistence of perturbations associated with small temporal scale which have high frequency components in their Fourier representation.

When all frequencies are excited equally as would be the case for uncorrelated white noise forcing of unit variance, i.e., $< \hat{p}_i(\omega_1) \; \hat{p}_j^*(\omega_2) > \; = \; \delta_{ij} \; \delta(\omega_1 - \omega_2)/2\pi$, the mean ensemble response variance weighted by $\mathbf{W}^f$, $\tilde{\mathbf{n}} = \mathbf{W}^f\mathbf{n}$, is

$$< |\tilde{\mathbf{n}}|^2 > = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega) \, d\omega, \tag{38}$$

in which the power spectrum is given by

$$F(\omega) = \mathrm{trace}\,(\mathbf{W}^f \; \mathbf{R} \; \mathbf{W}^i \; \mathbf{W}^{iT} \; \mathbf{R}^\dagger \; \mathbf{W}^{fT}) \, , \tag{39}$$

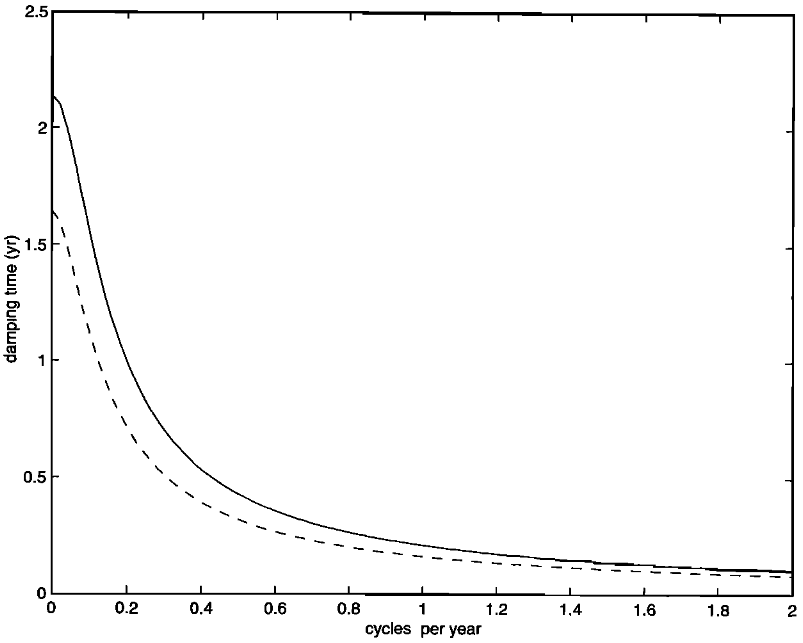


Figure 10. The damping time of troposphere-averaged CH_3Br number density (years) as a function of the frequency of the forcing of CH_3Br at the ground (cycles per year). Dashed curve shows damping time with only diffusion included. Continuous curve shows damping time with both advection and diffusion included.

where $\dagger$ denotes the hermitian transpose and $< |\tilde{n}|^2 >$ denotes the stationary ensemble variance. For forcing at the ground the power spectrum of mean CH_3Br mixing ratio variance over the troposphere (0 – 10 km) is shown in Figure 11 for forcing fluctuations with standard deviation 1 ppt per 12 min, as was used in the example shown in Figure 6.

7. Conclusions

A central problem in chemical dynamics is to determine the response of chemical equilibria to perturbation. This problem arises, for instance, in connection with assessing the impact of natural or anthropogenic sources. Traditional methods for addressing this problem include estimation of lifetimes from reservoir turnover times at equilibrium and calculation of modal decay rates. However, these methods may not be adequate to comprehensively analyze the response of chemical equilibria to perturbation particularly in cases in which these systems are nonnormal. Properties of the solutions of nonnormal systems are most effectively revealed by analysis of the system propagator. Methods based on the system propagator generalize directly to time-dependent problems while methods based on eigenanalysis do not generalize, because eigenmodes are not defined for time-varying systems. Among the general properties obtained using the methods of nonnormal system analysis which were illustrated using a model system in this work are bounds on optimal growth in the L_1 norm (sum of the absolute values of concentrations) and integral measure (absolute value of the sum of concentrations), the resulting growth as a function of time, the perturbation producing the optimal growth, the time period after which asymptotic decay rates are obtained, the response of the system to stochastic forcing applied impulsively and continually, and measures of timescales.

Because the perturbation chemical equilibrium system is linear and the appropriate measures are also linear, the optimal perturbation is a delta function corresponding to the introduction of a disturbance in the concentration of a single species at a single spatial location. At least conceptually, this greatly simplifies the problem of determining the optimal, because only a number of points equal to the dimension of the system are candidate optimals, while to determine the optimal in the L_2 norm, which is nonzero at all levels, it is necessary to search all vectors $\mathbf{n}$ in R^N satisfying $||\mathbf{n}||_2 = 1$, where N is the dimension of the system, which is a far wider space of candidate optimals. Remarkably, in the linear measures appropriate for chemistry a far greater simplification is possible: the optimal can be determined by inspection of the components of a single vector obtained from integration of the transposed system matrix.

The optimal excitation of a mode was found to be a delta function at the location corresponding to the maximum absolute value of the bi-orthogonal of the mode rather than of the mode itself. If the mode excited is

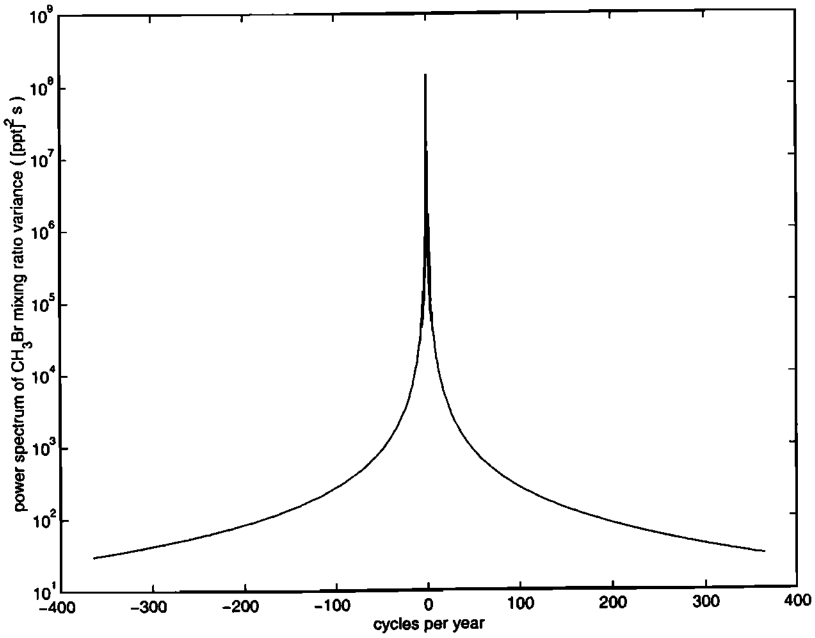


Figure 11. Power spectrum of CH_3Br mixing ratio variance averaged over the troposphere (0 – 10 km) resulting from white noise forcing of CH_3Br at the ground. The frequencies shown correspond to periods longer than a day. The peak response occurs for stationary forcing, and the response falls off rapidly with increasing frequency. Vertical advection is included.

the least damped mode, this corresponds to excitation of the most persistent perturbation. The excitation of the mode by placing a delta function initial concentration at the location where the absolute value of the biorthogonal is maximum leads to excitation of the mode at an amplitude which may far exceed the magnitude that would have been obtained if the mode itself were introduced initially (in the simple CH_3/Br_y example this led to a twofold increase of the amplitude of the least damped mode).

A single timescale for perturbations to a chemical equilibrium system with more than one degree of freedom does not exist, and only in the case that a single mode is excited is a single timescale obtained, and the timescale associated with the decay rate of the least damped mode is appropriate only in the asymptotic limit, assuming that a spectrum of modes is excited. Timescales based on reservoir turnover times and on relaxation from equilibrium depend on the spatial distribution of the species at equilibrium and the nonnormality of the system matrix. Moreover, reservoir timescales and timescales based on relaxation from equilibrium depend on the nonnormality of the operator, which in turn depends on the variable chosen and the measure used. Timescales obtained from observations of fluctuation variance require knowledge of the spatial distribution of the forcing and are, in general, short compared to the reservoir and relaxation from equilibrium timescales due to the excitation of highly damped high-frequency perturbations.

Appendix A: Optimal Growth and Optimal Perturbations in the Integral Measure

Consider the $N \times N$ propagator matrix $\mathbf{\Phi}(t)$ and a vector of number densities $\mathbf{n} = \sum_{i=1}^{N} \gamma_i \, \mathbf{e}^i$, where $\mathbf{e}^i$ is the canonical basis column vector with elements $e^i_j = \delta_{ij}$, and $\sum_{i=1}^{n} |\gamma_i| = 1$ so that $||\mathbf{n}||_1 = 1$. We have

$$\begin{aligned} \left\lfloor \mathbf{\Phi}(t)\mathbf{n} \right\rfloor &= \left| \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_j \; \Phi_{ij}(t) \right| \\ &\le \sum_{j=1}^{N} |\gamma_j| \left| \sum_{i=1}^{N} \Phi_{ij}(t) \right| \\ &\le \max_j \left[\left| \sum_{i=1}^{N} \Phi_{ij}(t) \right| \right] , \qquad \text{(A1)} \end{aligned}$$

where the last inequality follows from the fact that each $|\gamma_i| \le 1$ and the maximum sum over j corresponds to the column of $\mathbf{\Phi}(t)$ with maximum absolute sum. The maximum value of $\left\lfloor \mathbf{\Phi}(t)\mathbf{n} \right\rfloor$ over all unit vectors $\mathbf{n}$ is attained for the canonical basis vector $\mathbf{n} = \mathbf{e}^j$, where j is the column of $\mathbf{\Phi}(t)$ with the maximum absolute sum. We have proven that the optimal growth in the integral measure of the propagator, $||\mathbf{\Phi}(t)||$, is the largest absolute column sum of $\mathbf{\Phi}(t)$ and that the optimal vector producing this growth is the canonical basis vector corresponding to this column.

Similar results hold if instead of the integral measure, we consider the L_1 norm. The maximum value of $||\mathbf{\Phi}(t)\mathbf{n}||_1$ over vectors $\mathbf{n}$ of unit norm is attained for the canonical basis vector $\mathbf{n} = \mathbf{e}^j$, where j is the column of $\mathbf{\Phi}(t)$ with the maximum sum of the absolute values of its elements. Additionally, the perturbation producing the minimal disturbance in L_1 (although not necessarily in integral measure) is a canonical basis vector corresponding to the minimum column sum of the absolute values of the propagator.

This theorem permits a rapid search for the optimal growth and its optimal excitation. It is only required to determine the column of the propagator with maximum absolute sum. The optimal perturbation will be a unit perturbation at the corresponding level.

Although linear measures are often most appropriate for measuring chemical perturbations, in order to fix ideas it is useful first to consider the case in which the chosen perturbation measure is the L_2 norm. Then the optimal growth at time t is the norm of the weighted propagator initial perturbations of unit euclidean magnitude:

$$||\mathbf{W}^f \, e^{\mathbf{A}t} \, \mathbf{W}^i||_2 \; = \; \max_{||\mathbf{n}||_2=1} \; ||\mathbf{W}^f \, e^{\mathbf{A}t} \, \mathbf{W}^i\mathbf{n}||_2 \; . \qquad \text{(A2)}$$

Optimal growth at t in the L_2 norm is found by singular-value decomposition of the propagator $\mathbf{W}^f e^{\mathbf{A}t}\mathbf{W}^i$, and by this means a complete set of orthogonal perturbations is obtained, which can be ordered according to growth [*Farrell and Ioannou*, 1996a]. Analysis of transient growth in the integral measure appropriate for chemistry is less complete than that based on the L_2 norm, because this measure is not derived from an inner product, and consequently there is no associated concept of orthogonality by which to separate perturbations so that the perturbations can be ordered in contribution to the vector norm.

The results on the optimal perturbation in the integral norm can be illustrated and related to the familiar properties of the L_2 norm by considering the action on unit vectors of the model 2 × 2 propagator:

$$\mathbf{\Phi}(t) \; = \; \begin{pmatrix} -1 & -2 \\ 4 & 4 \end{pmatrix} \; . \qquad \text{(A3)}$$

In the L_2 norm the locus of the initial vectors of unit norm $||\mathbf{n}||_2 = 1$ is a unit circle in the (n_1, n_2) plane which is distorted into an ellipse by the action of the propagator matrix as illustrated in Figure 12. The length of the major semi-axis of this ellipse OB' gives the optimal growth, and the vector $\mathbf{OB}$ that is mapped to it is the optimal vector producing this growth. For this specific example the L_2 norm of the matrix is 6.05.

Consider now the unit vectors in the L_1 norm. The locus of initial vectors of unit norm $||\mathbf{n}||_1 = 1$ is the inscribed square which intersects the unit circle at the coordinate axes. This square is distorted into a parallelogram inscribed in the ellipse by the action of the propagator. We know from the above theorem that the optimal perturbation is located at one of the vertices of the inscribed square. A check of the vertices of the inscribed square reveals that the vertex C, which is mapped to C', is the optimal perturbation in the L_1 norm, and the optimal growth is the L_1 norm of $\mathbf{OC}'$, which is $||\mathbf{OC}'||_1 = 6$. The maximum growth in the integral measure is also achieved at one of the vertices, and a check of the vertices of the inscribed square reveals that this optimal perturbation is A, which is mapped to A'. The optimal growth is the integral measure of $\mathbf{OA}'$, which is $[\mathbf{OA}'] = 3$. Because the elements of the propagator $\mathbf{\Phi}$ have mixed signs the maximum absolute sum of the columns (the integral measure) differs from the sum of the absolute values of the columns (the L_1 norm). Note that in the L_1 norm and in the integral measure a vertex of the unit simplex is the optimal perturbation, and in the L_1 norm a vertex is also the perturbation resulting in minimal growth at time t. However, this property that the perturbation of minimal growth is at a vertex is not shared by the integral measure in which because of cancelation that may occur if the propagator has entries of both signs, a perturbation with zero growth can exist which is not associated with a vertex of the unit simplex.

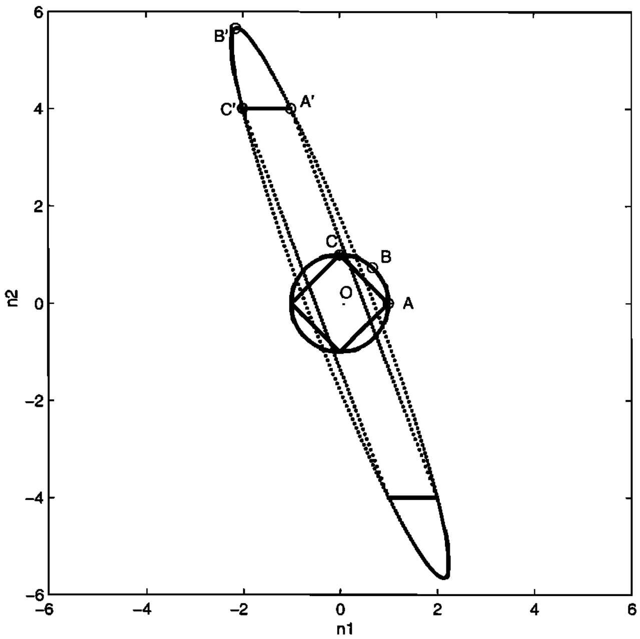


Figure 12. The result of applying the model 2 × 2 propagator (equation (A3)) to unit vectors in L_2 (the circle) and to unit vectors in L_1 (inscribed square). The optimal vector in L_2 is $\mathbf{OB}$, which is mapped to the evolved optimal coincident with the semimajor axis $\mathbf{OB}'$ of length 6.05, leading to $||\mathbf{\Phi}||_2 \; = \; \mathbf{6.05}$. The optimal vector in L_1 is $\mathbf{OC}$, and the L_1 norm of the evolved optimal, $\mathbf{OC}'$, is $||\mathbf{\Phi}||_1 \; = \; \mathbf{6}$. The optimal vector in integral measure is $\mathbf{OA}$, and it is mapped to the evolved optimal $\mathbf{OA}'$ with integral measure $[\mathbf{\Phi}] = \mathbf{3}$.

Appendix B: Optimal L_1 Excitations in the Asymptotic Limits

Consider first the $t \to \infty$ limit. We anticipate that every perturbation will assume the structure of the least damped mode in this limit. A similarity transformation diagonalizes the propagator: $e^{\mathbf{A}t} = \mathbf{U}\,\mathbf{D}\,\mathbf{U}^{-1}$, where $\mathbf{U}$ is the column matrix of the eigenvectors ordered according to $\text{Re}(\lambda)$, where λ is the eigenvalue of the corresponding eigenvector, and $\mathbf{D}$ is the diagonal matrix with elements $e^{\lambda t}$. For large time the first diagonal element of $\mathbf{D}$ will exponentially dominate, so that with exponential accuracy we obtain

$$||e^{\mathbf{A}t}|| = e^{\mathrm{Re}(\lambda_1)\mathrm{t}} \max_j \left[\left| \left(U^{-1}\right)_{1j} \right| \right] \left| \sum_i^N U_{i1} \right| , \qquad \text{(B1)}$$

where λ_1 is the eigenvalue of the least damped mode, U_{i1} denotes an element of the eigenvector of the least damped mode which we wish to optimally excite, and $\left[\left(U^{-1}\right)^T\right]_{j1}$ are the elements of its bi-orthogonal. The optimal excitation is seen to occur for a delta function at the location of the maximum absolute amplitude of the bi-orthogonal rather than at the maximum of the mode itself as might have been anticipated. This result contrasts with the optimal excitation of the least damped mode in the L_2 norm. In that case the bi-orthogonal of the least damped mode itself optimally excites the mode [*Farrell and Ioannou*, 1996a]. If eigenanalysis of $\mathbf{A}$ is not feasible, the optimal excitation of the least damped mode can be identified as the dominant vector in the limit $t \to \infty$ in the transposed system, which is the bi-orthogonal of the least damped mode.

In the $t \to 0$ limit the appropriate problem is to determine the initial perturbation, $\mathbf{n}$, of unit L_1 norm that leads to maximum instantaneous growth rate in the integral measure. This is given by

$$g = \lim_{t \to 0} \frac{||e^{\mathbf{A}t}|| - 1}{t} = \lim_{t \to 0} \frac{||\mathbf{I} + \mathbf{A}t|| - 1}{t}, \qquad \text{(B2)}$$

the second equality is obtained by Taylor expanding the propagator. The maximum growth rate is, in general, the maximum absolute sum obtained by a column of $\mathbf{A}$, and the perturbation that produces it is the canonical basis vector at the column of $\mathbf{A}$ corresponding to this maximum column sum (note that this is different from $||\mathbf{A}||_1$, which is the maximum obtained by a sum of absolute values of the columns of $\mathbf{A}$).

We may also obtain the perturbation that instantaneously grows most rapidly in the L_1 norm. An analysis similar to that above leads to the result that this maximum growth rate is the maximum sum of the columns of the matrix with elements

$$C_{ij} = \begin{cases} |A_{ij}| & i \neq j \\ A_{ij} & i = j, \end{cases} \qquad \text{(B3)}$$

and the perturbation that produces it is the canonical basis vector that corresponds to the column of $\mathbf{C}$ with the maximum column sum (note that this is again different from $||\mathbf{A}||_1$, unless all elements of $\mathbf{A}$ are positive). In the L_1 norm we can also determine the perturbation that leads to the minimum instantaneous growth rate: this perturbation is corresponding to the column of $\mathbf{C}$ with minimum column sum of absolute values.

Appendix C: Mean and Variance of the Statistical Steady State

C1. Continual Stochastic Forcing

Consider the response to continual stochastic forcing. Assuming $n_i(0) = 0$, the forced response at time t is

$$n_i(t) = \int_0^t e_{ij}^{\mathbf{A}(t-s)}\, W_{jk}\, p_k(s)\, ds , \qquad \text{(C1)}$$

where the propagator is $e^{\mathbf{A}t}$, $\mathbf{W}$ is the structure matrix of the forcing, and the vector forcings are considered to be white noise in space and time, i.e., $< p_k(t) p_l(s) > = \delta_{kl}\delta(t-s)$. The central limit theorem implies that the distribution of the evolved vector $n_i(t)$ approaches normal with mean at time t:

$$\bar{n}_i(t) = \left(\int_0^t e_{ij}^{\mathbf{A}(t-s)}\, W_{jk}\, ds \right) \bar{p}_k , \qquad \text{(C2)}$$

which as $t \to \infty$ becomes

$$\bar{n}_i(\infty) = -\, A_{ij}^{-1}\, W_{jk}\, \bar{p}_k . \qquad \text{(C3)}$$

The variance of $n_i(t)$ at each level can be obtained from the matrix

$$\mathbf{C}^t = \int_0^t e^{\mathbf{A}(t-s)}\, \mathbf{W}^i\, \mathbf{W}^{iT}\, e^{\mathbf{A}^T(t-s)}\, ds, \qquad \text{(C4)}$$

which satisfies the Lyapunov equation [*Farrell and Ioannou*, 1996a]

$$\mathbf{A}\,\mathbf{C}^\infty + \mathbf{C}^\infty\,\mathbf{A}^T = -\mathbf{W}^i\,\mathbf{W}^{iT} , \qquad \text{(C5)}$$

asymptotically as $t \to \infty$. The long time variance of n_i is C_{ii}^∞. Knowing the mean and the variance of n_i permits calculation of all statistical moments. An equivalent alternative view proceeds from Fourier analysis of the dynamical system to obtain the stochastic frequency response [*Farrell and Ioannou*, 1996a].

C2. Response to Impulsive Introduction of an Ensemble of Initial Perturbations

Consider now the expected growth of an ensemble of initial perturbations, $\mathbf{n}(0)$. The central limit theorem implies that the distribution of the evolved vector $n_i(t) = \sum_{j=1}^N \Phi_{ij}(t)\, n_j(0)$ approaches normal with mean:

$$\bar{n}_i(0) = \sum_{j=1}^N \Phi_{ij}(t)\, \bar{n}_j(0), \qquad \text{(C6)}$$

where $\overline{n}_j(0)$ is the mean of the ensemble of the initial perturbations at the jth grid and $\mathbf{\Phi}(t)$ is the propagator at time t. The variance at each level i is given by

$$\sigma_i^2 = \sum_{j=1}^{N} \Phi_{ij}^2(t) \ < n_j^2(0) >, \qquad \text{(C7)}$$

where $< n_j^2(0) >$ are the variances of the initial perturbations. Knowing the mean and the variance of $n_i(t)$ permits calculation of the expected value of $< |n_i(t)| >$ in closed form:

$$\begin{aligned} < |n_i(t)| > \ &= \ \sqrt{\frac{2}{\pi}}\, \sigma_i \, \exp\left(- \frac{\overline{n}_i(t)^2}{2\sigma_i^2} \right) \ + \\ &+ \ \overline{n}_i(t) \, \mathrm{erf}\left(\frac{\overline{n}_i(t)}{\sqrt{2}\sigma_i} \right) \, , \qquad \text{(C8)} \end{aligned}$$

and consequently the expected growth of an ensemble of initial conditions is given in the L_1 norm by

$$< ||\mathbf{n}(t)||_1 > \ = \ \sum_{i=1}^{N} \ < |n_i(t)| > \ . \qquad \text{(C9)}$$

Appendix D: Effect of Nonnormality on Reservoir Turnover Time

We show that the turnover time defined in (21) depends on the nonnormality of **A**. Consider the simple 2 × 2 system matrix:

$$\mathbf{A}(t) = \begin{pmatrix} -0.1 & 10 \\ 0 & -0.2 \end{pmatrix} , \qquad \text{(D1)}$$

representing in a schematic form the processes of advection and dissipation. This matrix is nonnormal and the reservoir turnover time for a source at the first grid is easily found to be 10, i.e., $|| - \mathbf{A}^{-1}[1,0]^T||_1 = 10$, while for a source at the second grid the reservoir turnover time is 505, i.e., $|| - \mathbf{A}^{-1}[0,1]^T||_1 = 505$. The corresponding turnover times for the diagonal normal matrix with the same eigenvalues are 10 and 5. So two systems with the same decay rates of their natural modes possesses vastly different reservoir turnover times. This demonstrates that the reservoir turnover times can be importantly affected by the nonnormality of **A**.

Acknowledgments. The authors wish to acknowledge discussions with Denise Mauzerall, Larry Horowitz, and Steve Wofsy.

References

Andrews, D. G., J. R. Holton, and C. B. Leovy, *Middle Atmospheric Dynamics*, Academic, San Diego, Calif., 1987.

Chameides, W. L., and E. M. Perdue, *Biogeochemical Cycles*, Oxford Univ. Press, New York, 1997.

Farrell, B. F., Optimal excitation of perturbations in viscous shear flow, *Phy. Fluids*, *31*, 2093-2102, 1988.

Farrell, B. F., and P. J. Ioannou, Generalized stability, part I, Autonomous operators, *J. Atmos. Sci.*, *53*, 2025-2040, 1996a.

Farrell, B. F., and P. J. Ioannou, Generalized stability, part II, Non autonomous operators, *J. Atmos. Sci.*, *53*, 2041-2053, 1996b.

Golub, G. H., and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, Md., 1996.

Khalil, M. A. K., and R. A. Rasmussen, Modeling chemical transports and mass balances in the atmosphere, in *Environmental Exposure From Chemicals, Vol. 2*, edited by W. B. Neely and G. E. Blau, CRC Press, Boca Raton, Fla., 1984.

Prather, M. J., Timescales in atmospheric chemistry: theory, GWPs for CH_4 and CO, and runaway growth, *Geophys. Res. Lett.*, *23*, 2597-2600, 1996.

Prather, M. J., Timescales in atmospheric chemistry: CH_3Br, the ocean, and ozone depletion potentials, *Global Biogeochem. Cycles*, *11*, 393-400, 1997.

Prather, M. J., Timescales in atmospheric chemistry: Coupled perturbations N_2O, NO_y, and O_3, *Science*, *279*, 1339-1341, 1998.

Trefethen, L. N., Pseudospectra of matrices, in *Numerical Analysis*, edited by D.F. Griffiths and G. A. Watson, pp. 234-266, Addison-Wesley-Longman, Reading, Mass., 1991.

B. Farrell, Department of Earth and Planetary Sciences, Harvard University, 29 Oxford Street, Cambridge, MA 02138. (farrell@io.harvard.edu)

P. Ioannou, Section of Astrophysics, Astronomy and Mechanics, Department of Physics, National and Capodistrian University of Athens, Zografos 15784, Athens, Greece. (pji@cc.uoa.gr)

(Received August 26, 1998; revised September 29, 1999; accepted September 30, 1999.)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9321-9333, APRIL 16, 2000

Influence of convection and biomass burning outflow on tropospheric chemistry over the tropical Pacific

Yuhang Wang,[1] Shaw C. Liu, Hongbin Yu, and Scott T. Sandholm
School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta

Tai-Yih Chen and Donald R. Blake
Department of Chemistry, University of California, Irvine

Abstract. Observations over the tropics from the Pacific Exploratory Mission-Tropics A Experiment are analyzed using a one-dimensional model with an explicit formulation for convective transport. Adopting tropical convective mass fluxes from a general circulation model (GCM) yields a large discrepancy between observed and simulated CH_3I concentrations. Observations of CH_3I imply the convective mass outflux to be more evenly distributed with altitude over the tropical ocean than suggested by the GCM. We find that using a uniform convective turnover lifetime of 20 days in the upper and middle troposphere enables the model to reproduce CH_3I observations. The model reproduces observed concentrations of H_2O_2 and CH_3OOH. Convective transport of CH_3OOH from the lower troposphere is estimated to account for 40-80% of CH_3OOH concentrations in the upper troposphere. Photolysis of CH_3OOH transported by convection more than doubles the primary HO_x source and increases OH concentrations and O_3 production by 10-50% and 0.4 ppbv d^{-1}, respectively, above 11 km. Its effect on the OH concentration and O_3 production integrated over the tropospheric column is, however, small. The effects of pollutant import from biomass burning regions are much more dominant. Using C_2H_2 as a tracer, we estimate that biomass burning outflow enhances O_3 concentrations, O_3 production, and concentrations of NO_x and OH by 60%, 45%, 75%, and 7%, respectively. The model overestimates HNO_3 concentrations by about a factor of 2 above 4 km for the upper one-third quantile of C_2H_2 data while it generally reproduces HNO_3 concentrations for the lower and middle one-third quantiles of C_2H_2 data.

1. Introduction

Convection is a dominant process for the vertical distribution of chemical species in the troposphere. This distribution process, unlike diffusion, is strongly direction oriented. Updrafts bring up air from the lower troposphere into the free troposphere, and compensating subsidence takes free tropospheric air back down into the lower troposphere. Over industrialized continents, convection leads to efficient export of pollutants into the free troposphere and greatly enhances photochemistry outside the polluted continental lower troposphere [*Gidel*, 1983; *Chatfield and Crutzen*, 1984; *Dickerson et al.*, 1987; *Pickering et al.*, 1988, 1992; *Jaeglé et al.*, 1998].

Over tropical continents, convection exports large amounts of pollutants from biomass burning [*Crutzen and Andreae*, 1990]. *Fishman et al.* [1990] showed widespread high tropospheric O_3 columns over the tropical South Atlantic and adjacent continents during the biomass burning season (September-November). Estimated tropospheric O_3 column in the region is a factor of 2 higher than over the tropical Pacific. During the Transport and Atmospheric Chemistry Near the Equatorial Atlantic (TRACE A) Experiment in September-October, 1992, elevated concentrations of O_3, NO_x ($NO+NO_2$), CO, and hydrocarbons from biomass burning emissions were observed at all altitudes [*Fishman et al.*, 1996]. Transport of these pollutants evidently influences chemistry in the remote tropical Pacific [*Jacob et al.*, 1996; *Schultz et al.*, 1999].

Recent works suggest that local convection in remote tropical regions also has a large impact on chemistry in the upper troposphere [*Prather and Jacob*, 1997; *Jaeglé et al.*, 1997; *Cohan et al.*, 1999]. This effect arises from the large gradient of chemical species between the lower and upper troposphere. Photochemistry driven by radical production from photolysis of O_3 and the subsequent reaction of O^1D+H_2O is active in the lower troposphere due in part to the abundance of H_2O. In the upper troposphere, photochemistry is considerably slower; radical sources, such as photolysis of acetone, become much more important than O^1D+H_2O [*Singh et al.*, 1995; *Arnold et al.*, 1997; *McKeen et al.*, 1997b; *Jaeglé et al.*, 1997; *Wennberg et al.*, 1998; *Müller and Brasseur*, 1999]. As a result, photochemically produced chemical species tend to have much higher concentrations in the lower than in the upper troposphere. *Prather and Jacob* [1997] noted the large vertical gradient of CH_3OOH in the tropics and suggested that convection of CH_3OOH is a dominant radical source in the tropical upper troposphere.

---

[1] Now at Department of Environmental Sciences, Rutgers University, New Brunswick, New Jersey.

Copyright 2000 by the American Geophysical Union.

Paper number 1999JD901127.
0148-0227/00/1999JD901127$09.00

It is, however, not clear to what extent the radical source from convected CH_3OOH affects O_3 and OH concentrations in the tropical troposphere and how this effect compares with that of biomass burning. Using a one-dimensional model, we analyze tropical DC-8 observations from the Pacific Exploratory Mission-Tropics A (PEM-Tropics A) Experiment in September-October, 1996 [*Hoell et al.*, 1999]. Previous photochemical analyses for PEM-Tropics A [e.g., *Schultz et al.*, 1999; *Cohan et al.*, 1999; *Crawford et al.*, 1999] have used point models, which cannot directly simulate the effect of convection. The vertical transport in our model, which includes an explicit treatment for convective transport, is constrained using the observations of CH_3I since the vertical distribution of CH_3I is a sensitive indicator for marine convection [*Davis et al.*, 1996b; *Cohan et al.*, 1999]. On the basis of observations of O_3, NO, CO, and hydrocarbons, we investigate if the model can simulate observed vertical profiles of other chemical species such as peroxides and use the model to compute the column budget of O_3. We will examine the effect of convection on chemistry. We will also use C_2H_2 as a tracer of outflow from biomass burning regions [*Blake et al.*, 1996, 1999; *Andreae et al.*, 1996] to examine the effect of this pollutant import and compare this effect with that of local convection.

2. Model Description

The model is based on the one-dimensional model by *Trainer et al.* [1987, 1991] and *McKeen et al.* [1997a]. We extend the vertical domain to 16 km, the altitude of the tropical tropopause. The vertical resolution of the model decreases from 10 m near the surface to 1 km at 12-km altitude, which is the flight ceiling of the DC-8 aircraft. The top two layers are equally spaced between 12 and 16 km. Chemical species with lifetimes longer than 10 min are transported vertically. The time step for transport and chemistry is 30 s. The model is run for 60 days to obtain a steady state solution.

The mass continuity equation for a one-dimensional model is

$$\frac{\partial n_i}{\partial z} = P_i + L_i + \frac{\partial \Phi_i}{\partial z} \ , \qquad (1)$$

where n_i is the concentration of species i, P_i is the chemical production rate, L_i is the loss rate by chemical reactions and in the lowest model layer by dry deposition, and Φ_i is the flux due to vertical transport. We consider two types of vertical transport, diffusion and convection. Diffusion, which is the only transport mechanism used by *Trainer et al.* [1987], represents mixing by random motion of air parcels. In contrast, convection represents direct transport of the lower tropospheric air into the free troposphere followed by compensating subsidence. The need for invoking convective transport in our model becomes apparent in the simulation of CH_3I in section 3. Including convective transport also allows us to explicitly simulate wet scavenging of soluble species in the model. Previous one-dimensional models usually invoke a pseudo first-order loss rate constant to account for wet scavenging of soluble species [e.g., *Logan et al.*, 1981].

The vertical flux Φ_i is the sum of diffusive and convective fluxes. The diffusive flux Φ_d is a function of the vertical gradient of the chemical species i and the eddy diffusion coefficient K_z,

$$\Phi_d = N K_z \frac{\partial (n_i / N)}{\partial z} \ , \qquad (2)$$

where N is the concentration of air. The value of K_z in the surface layer (lower than 80 m in altitude) is calculated from similarity theory [*Trainer et al.*, 1987]. The value of K_z above the surface layer will be specified. Convective mass fluxes in the model are from the Goddard Institute for Space Studies (GISS) general circulation model (GCM) (version II') as previously used by *Prather and Jacob* [1997]. We will discuss the model vertical transport constrained by CH_3I observations in section 3.

The chemical mechanisms in the model are updated from the compilations by *DeMore et al.* [1997] and *Atkinson et al.* [1997]. Quantum yields and cross sections for photolysis of acetone are from *Gierczak et al.* [1998]. The reaction rate constants of OH with HNO_3 and NO_2 are updated from *Brown et al.* [1999a, b]. The model does not include N_2O_5 hydrolysis on aerosols [*Schultz et al.*, 2000]. The total ozone column is specified to be 268 Dobson units (DU), the average measured for the tropical region during PEM-Tropics A by the Total Ozone Mapping Spectrometer. A surface albedo of 0.1 is specified for the ocean. We compared model computed photolysis rates of $J(O^1D)$ and $J(NO_2)$ with observations for solar zenith angle <50°, when the value of $J(NO_2)$ is largely invariable to solar zenith angle. The model averages agree to within ±12% with mean and median values of the observations binned in 1-km intervals. Dry deposition velocities to water are specified to be 1 cm s^{-1} for soluble species HNO_3 and H_2O_2 and less than 0.1 cm s^{-1} for other species [*Wang et al.*, 1998a].

3. Vertical Transport Constrained by CH_3I Observations

We use observed CH_3I concentrations to adjust the parameters of our vertical transport scheme. Methyl iodide is emitted largely from the ocean [*Andreae*, 1990; *Bates et al.*, 1992; *Happell and Wallace*, 1996]. It is lost primarily by photolysis. Its photochemical lifetime is about 3 days near the surface and 2 days in the upper troposphere. Its vertical distribution therefore depends critically on the vertical transport of CH_3I from the marine lower troposphere. Furthermore, the detection limit of CH_3I during PEM-Tropics A is low (0.01 parts per trillion by volume (pptv)) [*Cohan et al.*, 1999], making CH_3I observations a valuable constraint on the model transport scheme. *Blake et al.* [1996] estimated that CH_3I emissions from biomass burning during TRACE-A are much smaller than the oceanic source. The vertical profiles of CH_3I compiled for the lower, middle, and upper one-third quantiles of C_2H_2, a tracer for biomass burning, also do not exhibit enhancements of CH_3I associated with upper quantiles of C_2H_2, suggesting that observed CH_3I originates primarily from the ocean. We specified the CH_3I concentration at 500-m altitude in the model to the observed value.

Figure 1 shows simulated CH_3I concentrations using different vertical transport schemes. The two diffusion-only simulations show the largest discrepancies. For these simulations we specified the values of K_z as 10 and 60 m s^{-2} above the surface layer, respectively. For comparison, values of K_z in the troposphere derived from measurements of ^{222}Rn in the eastern United States vary from 40 m^2 s^{-1} in summer to 20 m^2 s^{-1} in

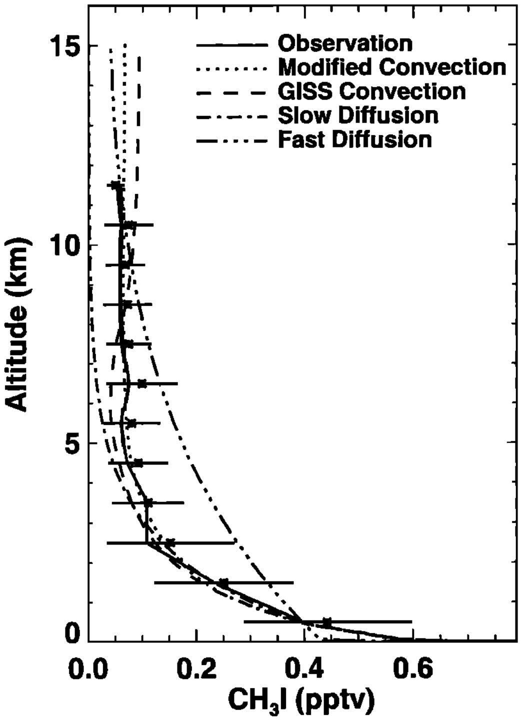


Figure 1. Comparison between observed and simulated vertical profiles of CH_3I concentrations. The observations over the tropical Pacific (20°S-20°N) are binned vertically in 1-km intervals. The solid line represents the median profile. Asterisks and horizontal bars are means and standard deviations, respectively, of the observations. Four model simulations using different vertical transport schemes are shown (see text for details). The concentration of CH_3I at 500 m is specified in the model as the observed median at 0-1 km.

spring and fall [*Liu et al.*, 1984]. In the slow diffusion case (K_z = 10 m s^{-2}) the simulation agrees well with the observations below 2 km but drastically underestimates CH_3I concentrations in the middle and upper troposphere. Adopting a faster diffusion coefficient (K_z = 60 m s^{-2}) elevates simulated CH_3I concentrations to observed values at 10-12 km but also leads to gross overestimates in the middle and lower troposphere. These large discrepancies reflect the dominance of tropical convective transport, which cannot be approximated by a diffusion scheme.

Table 1 shows the convective mass outflux over the equatorial region (12°S-12°N) from the GISS GCM, version II' [*Rind and Lerner*, 1996]. The statistics of nonentraining, deep convection have been used by *Prather and Jacob* [1997]. The convective mass outflux below 10 km is significantly larger in Table 1 than reported by *Prather and Jacob* [1997] due to the addition of entraining convection in the statistics. Figure 1 shows the simulation of CH_3I concentrations based on these convective statistics. The value of K_z is specified as 10 m s^{-2} above the surface layer to reproduce the observed CH_3I vertical gradient in the lower troposphere. This simulation is much closer to the observations compared with diffusion-only simulations. However, it overestimates CH_3I concentrations at 9-12 km but underestimates at 4-7 km, reflecting the "C"-shaped profile of the GISS convective statistics. Similar discrepancies were found in a global simulation of CH_3I driven by assimilated meteorological data from the Goddard Earth Observation System Data Assimilation System (M. Schultz, personal communication, 1999).

Observations of CH_3I in the tropics therefore argue for a more uniformly distributed profile of the convective mass outflux with altitude over the tropical ocean than suggested by the GISS model. *Smyth et al.* [1996] also found that mixing of CO and hydrocarbons from the continental boundary layer into the middle troposphere is as efficient as mixing into the upper troposphere in the tropical western Pacific. We redistributed the GISS convective outflux of deep convection (>3-km altitude) evenly by air density (Table 1). The difference in mass outflux above 3 km reflects the difference in air mass among different layers. The resulting CH_3I concentrations in the model agree well with observations (Figure 1). The redistributed convective mass outflux, which yields a uniform convective turnover lifetime of 20 days above 3 km, will be used in the model to examine the effect of convection on chemistry in the tropics (section 5). Our results cannot rule out the possibility of greater convective outflux above 12 km compared with lower altitudes as indicated by the GISS model, because observations of CH_3I are unavailable above 12 km during PEM-Tropics A.

4. Influence of Biomass Burning Outflow

Biomass burning took place in Australia, southern Africa, and South America during the PEM-Tropics A period [*Olson et al.*, 1999]. Its strong influence was reflected in the observations over the tropical Pacific [*Blake et al.*, 1999; *Schultz et al.*, 1999; *Talbot et al.*, 1999]. A good tracer of tropical biomass burning is C_2H_2 [*Blake et al.*, 1996, 1999; *Andreae et al.*, 1996]; almost all C_2H_2 data between 0° and 20°N were below 30 pptv, whereas observations in the southern tropics often exceeded this level. We group observations in the tropics (20°S-20°N) on the basis of concurrent measurements of C_2H_2 to analyze the effect of biomass burning.

Three vertical profiles of observed chemical species (binned in 1-km intervals) are compiled corresponding to the

Table 1. Convective Mass Outflux

Altitude Range, km	GISS Flux[a], g m^{-2} s^{-1}	Modified Flux[b], g m^{-2} s^{-1}
10-16	1.6	0.88
7.5-10	0.8	0.76
5-7.5	0	0.90
3-5	1.1	0.96
1.5-3	2.2	2.2
0.7-1.5	-3.1	-3.1
0-0.7	-2.6	-2.6

[a] Convective statistics over the equatorial region (12°S-12°N) in the GISS II' general circulation model between May and August (M. Prather, personal communication, 1999). Positive values are mass outfluxes of convection; negative values reflect convergence in the lower troposphere feeding the convection. The upward convective mass transport is compensated for by subsidence as required by mass balance.

[b] Redistributed uniformly above 3 km by air density to reproduce observed CH_3I concentrations (Figure 1).

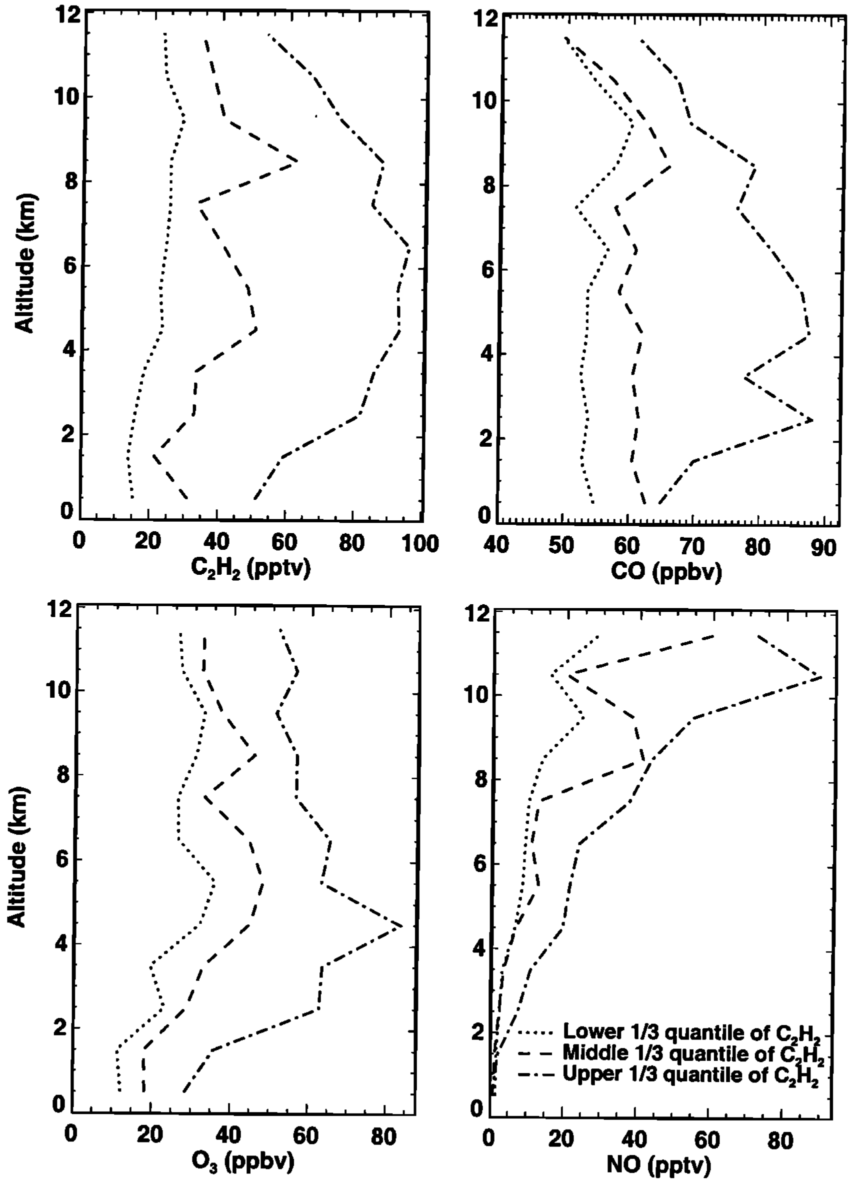


Figure 2. Observed median vertical profiles of C_2H_2, CO, O_3, and NO for three equally divided quantiles of C_2H_2 over the tropical Pacific. Observations are binned vertically in 1-km intervals.

lower, middle, and upper one-third quantiles of C_2H_2 concentrations. This simple grouping separates air masses significantly influenced by biomass burning outflow with high C_2H_2, CO, O_3, and NO concentrations from those with much lower concentrations (Figure 2). The median concentration of C_2H_2 increases from about 20 pptv in the lower one-third quantile to 60-80 pptv in the upper one-third quantile. The chemical lifetime of C_2H_2 against OH oxidation is only 1-2 weeks in the tropics. The factor of 3-4 enhancement of C_2H_2 concentrations from the lower to the upper one-third quantile therefore indicates strong influence of biomass burning outflow in the region. Another biomass burning tracer, CO, shows an increase of 20-80% from the lower to the upper one-third quantile of C_2H_2. This increase is smaller because of background CO concentrations from CH_4 oxidation [e.g., *Logan et al.*, 1981]. The small CO vertical gradient is due to its much longer lifetime (about 1 month in the lower troposphere and over 2 months in the upper troposphere) relative to convective turnover, which we estimate to be about 20 days (section 3). If CH_4 oxidation were its sole source, the CO concentration would be about 40 ppbv in the tropics. The enhancement of CO over a background of 40 ppbv shows large increases from the lower to the upper one-third quantile of C_2H_2, which is consistent with the enhancement of C_2H_2. Biomass burning also emits large amounts of NO that can lead to enhanced O_3 concentrations from the increasing in situ production [e.g., *Crutzen and Andreae*, 1990]. Concentrations of NO in biomass burning outflow may be further enhanced by lightning production [e.g., *Wang et al.*, 1998b]. Observed concentrations of O_3 and NO concentrations show corresponding increases with C_2H_2 (Figure 2). Column O_3 and NO_x concentrations (0-12 km) for different quantiles of C_2H_2 are listed in Table 2.

Table 2. Column O_3 Budget and OH and NO_x Concentrations

One-Third Quantile of C_2H_2	O_3[a]	$P(O_3)$[a]	$L(O_3)$[a]	OH[b]	NO_x[c]
Lower	14	1.4	3.7	1.4	2.0
Middle	20	1.7	4.5	1.5	3.1
Upper	34	3.1	5.5	1.6	5.4

Integrated over the air column of 0-12 km (20°S-20°N). Model values are 24-hour averages.

[a] Model-computed chemical production of O_3 from reactions of NO and peroxy radicals and loss from the reaction of O^1D+H_2O and reactions of O_3 with OH and HO_2 [*Davis et al.*, 1996a]. Column ozone is in Dobson units. Column production ($P(O_3)$) and loss ($L(O_3)$) of O_3 are in 10^{11} molecules cm^{-2} s^{-1}.

[b] Mass weighted column mean OH concentrations (in 10^6 molecules cm^{-3}) computed using the one-dimensional model.

[c] Concentrations of NO_x are computed using the one-dimensional model constrained to match observed daytime NO concentrations shown in Figure 2. Column NO_x is in 10^{14} molecules cm^{-2}.

5. Model Simulations: Effects of Convection and Biomass Burning Outflow

We constrain the one-dimensional model by observed median concentrations of O_3, NO, CO, nonmethane hydrocarbons (NMHCs), and H_2O for the lower, middle, and upper one-third quantiles of C_2H_2. Concentrations of NO_x are constrained in the model such that simulated daytime mean NO concentrations match the observed NO profiles. Among NMHCs, only long-lived C_2H_6, C_3H_8, and C_2H_2 were observed in significant amounts [*Blake et al.*, 1999]. Acetone was not measured during PEM-Tropics A. Observations during the Pacific Exploratory Mission (PEM-West B) show acetone concentrations above 200 pptv and below 400 pptv in the tropics [*Singh et al.*, 1995]. We therefore specify acetone concentrations of 200 and 400 pptv for data groups of the lower and upper one-third quantiles of C_2H_2, respectively. We specify for the middle one-third quantile of C_2H_2 a concentration of 300 pptv based on a global three-dimensional model simulation by *Wang et al.* [1998a, b]. The concentration of CH_4 is specified at 1.7 ppmv. The DC-8 flight ceiling was at about 12 km; we assume constant mixing ratios above 12 km to be the observed medians at 11-12 km for most species except for water vapor, which decreases with altitude to 6 ppmv at the tropopause.

5.1. Hydrogen Oxides

Concentrations of OH or HO_2 radicals were not measured on the DC-8 during PEM-Tropics A. We evaluate the model simulations using observations of the reservoirs of these radicals, H_2O_2 and CH_3OOH. In situ chemistry generally plays a central role in determining peroxide concentrations since their lifetimes in the tropics are less than 1 day in the lower troposphere and about 2 days in the upper troposphere. An exception is in the upper troposphere, where convective transport can greatly change their concentrations. In section 3 we have estimated a convective turnover lifetime of 20 days for the ventilation of the upper troposphere. Although this timescale is considerably longer than the lifetimes of peroxides, the peroxide concentrations are about a factor of 10 higher in the lower than in the upper troposphere (Figure 3). As a result, convective transport of air masses from the lower troposphere may provide a source of peroxides similar to in situ production. This effect does not apply to H_2O_2 since H_2O_2 is soluble, whereas the solubility of CH_3OOH is very low [*Lind and Kok*, 1986]. *Cohan et al.* [1999] estimated that 50-70% of convected H_2O_2 is scavenged, suggesting that convection has a small impact on upper tropospheric H_2O_2 concentrations.

Figure 3 compares observed and simulated H_2O_2 concentrations. We conducted simulations with wet scavenging coefficients of 50% and 100% for H_2O_2. A 100% scavenging coefficient corresponds to a complete removal of H_2O_2 from the convective air masses. The lower H_2O_2 concentrations simulated with a scavenging coefficient of 100% are in closer agreement with observations in the upper troposphere. The difference between the two simulations is small in the lower and middle troposphere. Thus we select the scavenging coefficient of 100% in our standard model.

Hydrogen peroxide is produced by the self-reaction of HO_2 and chemically destroyed by photolysis and the reaction with OH. The HO_2 self-reaction is the dominant pathway for HO_x (OH + peroxy radicals) loss under low NO_x conditions in the lower troposphere [*Kleinman*, 1994]. The HO_x loss is balanced by the HO_x production, which is largely photolysis of O_3 followed by the reaction of O^1D+H_2O in the lower and middle troposphere [*Logan et al.*, 1981]. Although O_3 concentrations increase by a factor of 2 from the lower to the upper one-third quantile of C_2H_2 (Figure 2), air masses with high C_2H_2 concentrations were much drier (Figure 4). The relative humidity for the lower one-third quantile C_2H_2 decreases from 80% in the lower troposphere to 15% in the upper troposphere. It is 50-300% higher than that for the upper one-third quantile of C_2H_2. The compensating effects of O_3 and H_2O on the source of HO_x are reflected in comparable H_2O_2 concentrations for the three quantiles of C_2H_2 simulated in the model.

Observed H_2O_2 concentrations are similar for the lower and middle one-third quantiles of C_2H_2 but are lower for the upper one-third quantile of C_2H_2. The lower H_2O_2 concentrations (in the lower and middle troposphere) are not reproduced by the model. Comparison of observed and simulated CH_3OOH concentrations (Figure 3) shows a more pronounced overestimate for the upper one-third quantile of C_2H_2. The CH_3OOH budget for 0°-30°S by *Schultz et al.* [1999] also shows larger chemical production of CH_3OOH than loss in the lower and middle troposphere. A possible explanation is that dry, high C_2H_2 concentration air masses, which originated most likely from the continents, have not reached chemical steady state with respect to peroxides and hence have lower peroxide concentrations.

Convective transport largely enhances CH_3OOH concentrations in the upper troposphere (Figure 3), as suggested by *Prather and Jacob* [1997]. Model results with convective

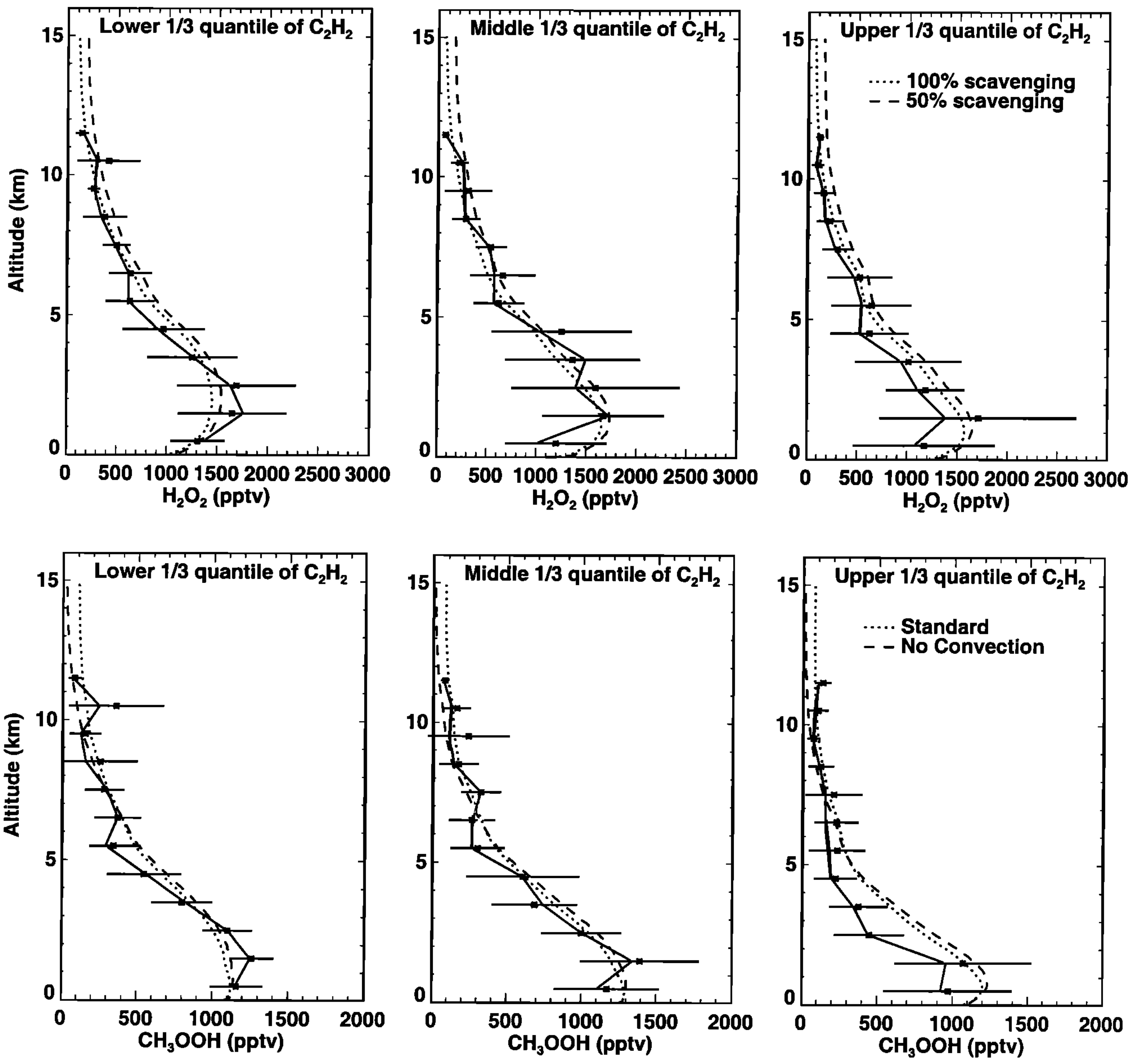


Figure 3. Comparison of observed and simulated vertical profiles of H_2O_2 and CH_3OOH for the lower, middle, and upper one-third quantiles of C_2H_2. Symbols for the observations are the same as in Figure 1. Model simulations of H_2O_2 concentrations are shown with scavenging coefficients of 100% and 50%. Model simulations of CH_3OOH concentrations are shown with and without convective transport.

transport agree better with the observations above 10 km than without convection. Figure 5 shows the fraction of CH_3OOH attributed to convective transport in our standard model. The fraction is large in the upper troposphere, increasing from 40% at 10 km up to 80% at 15 km. Below 6 km the fraction drops to less than 20%. In this region the increase of CH_3OOH from convective transport is not enough to offset the decrease due to subsidence of air with lower CH_3OOH concentrations, resulting in lower CH_3OOH concentrations compared with the simulation without convection (Figure 3).

Photolysis of CH_3OOH convected from the lower troposphere provides a significant primary HO_x source that drives chemistry faster in the upper troposphere. Figure 6 compares primary HO_x sources, O^1D+H_2O, photolysis of acetone, and photolysis of CH_3OOH and CH_2O transported from the lower troposphere by convection (for the middle one-third quantile of C_2H_2). Photolysis of O_3 followed by the reaction of O^1D+H_2O dominates the primary HO_x sources up to 10 km. The rapid decrease of this source with altitude reflects decreasing H_2O levels (Figure 4). The HO_x sources from photolysis of acetone and convected CH_3OOH are less variable with altitude and surpass that from O^1D+H_2O above 11 km. Convective transport of CH_3OOH more than doubles the total primary source of HO_x in the upper troposphere. The source from photolysis of convected CH_2O is a factor of 5-10 smaller than that from convected CH_3OOH.

The total HO_x production is larger than the sum of its primary sources because of photolysis of H_2O_2 and CH_3OOH

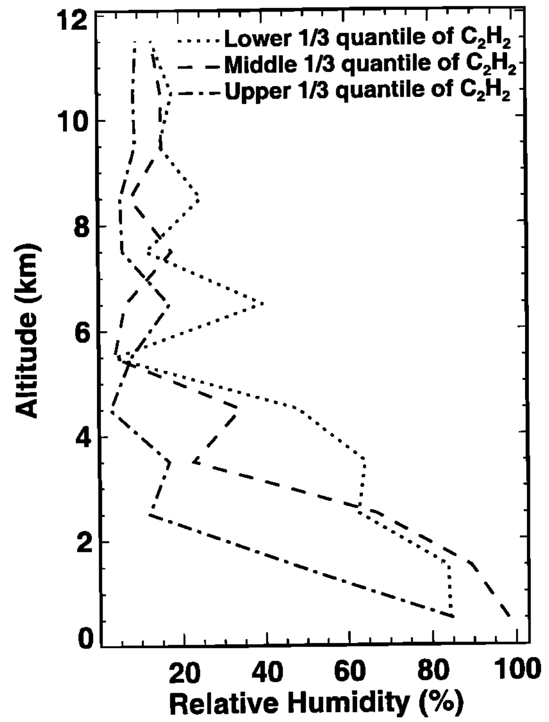


Figure 4. Same as Figure 2 but for relative humidity (over water). Water vapor concentrations measured by the differential absorption CO measurement (DACOM) instrument are used.

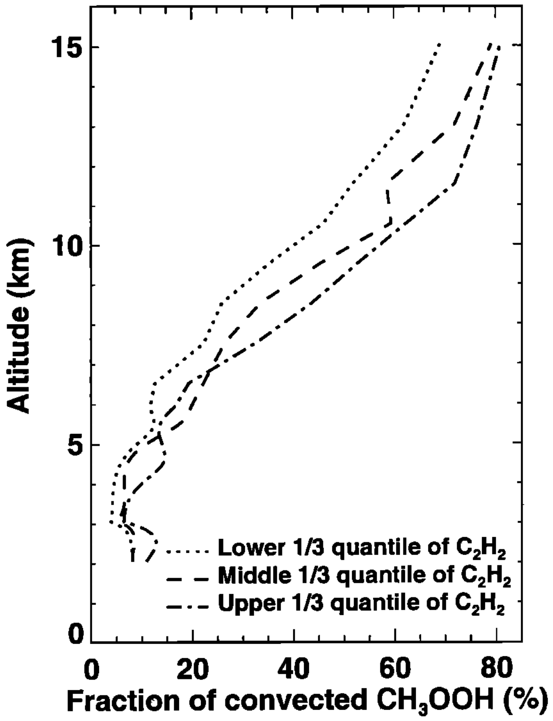


Figure 5. Fraction of CH_3OOH attributed to convective transport as a function of altitude for the lower, middle, and upper one-third quantiles of C_2H_2.

produced in situ and photolysis of CH_2O from CH_4 oxidation. Figure 7 compares the total HO_x production with and without convection. The increase of total HO_x production due to convective transport of CH_3OOH is apparent above 10 km by up to a factor of 3. Below 8 km, HO_x production with convection tends to be lower than without convection due to lower H_2O_2 and CH_3OOH concentrations in the convective case (Figure 3). The concentration of OH shows corresponding changes but with smaller magnitudes (Figure 7); the maximum increase is less than 50% in the upper troposphere. The increase of HO_x production due to convection of CH_3OOH has a negligible effect on the column mean OH concentration.

5.2. Ozone

The additional HO_x source from CH_3OOH transported by convection enhances in situ O_3 production in the upper troposphere [*Prather and Jacob*, 1997; *Jaeglé et al.*, 1997]. Figure 8 compares the O_3 production rate between simulations with and without convection for the middle one-third quantile of C_2H_2. Ozone is produced by reactions of NO and peroxy radicals. The production rate increases by about 0.4 ppbv d^{-1} above 11 km. Integrated over the tropospheric column (0-16 km), the production rate of O_3 increases by only 4% from 1.88×10^{11} to 1.96×10^{11} molecules cm^{-2} s^{-1}.

The effect of biomass burning outflow is much larger in comparison. Observed ozone columns of 0-12 km are 14, 20, and 34 DU for the lower, middle, and upper one-third quantiles, respectively, of C_2H_2 (Table 2). Assuming the lower one-third quantile of C_2H_2 data are not significantly influenced by outflow from biomass burning regions, we estimate an increase of 60% or 9 DU due to enhanced O_3 concentrations in

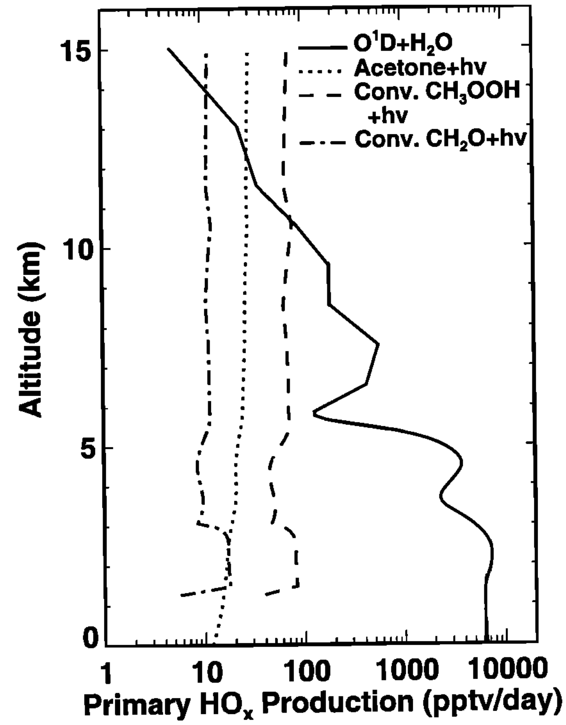


Figure 6. Simulated primary sources of HO_x, O^1D+H_2O, photolysis of acetone, and photolysis of CH_3OOH and CH_2O transported from the lower troposphere by convection, as a function of altitude for the middle one-third quantile of C_2H_2. The HO_x yield of CH_2O photolysis is computed on line.

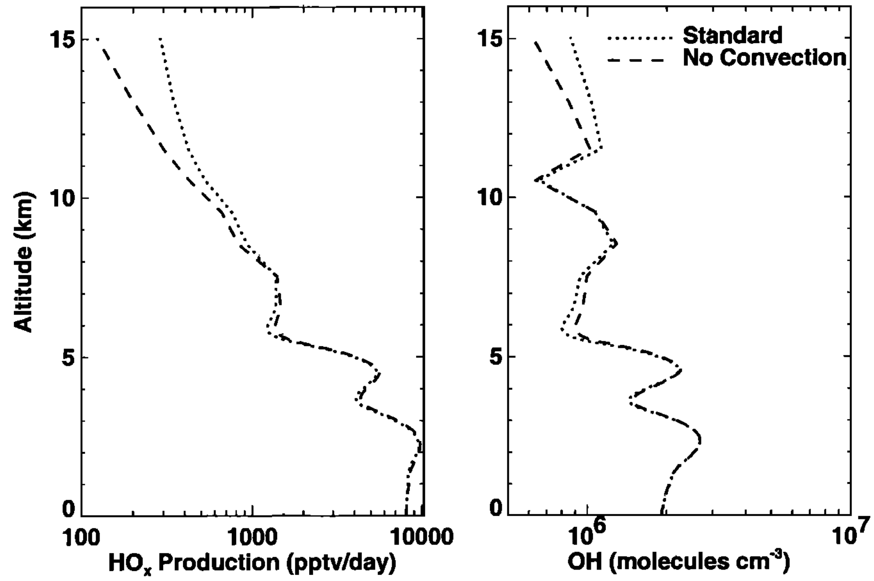


Figure 7. Simulated total HO_x production and OH concentrations as a function of altitude for the middle one-third quantile of C_2H_2 with and without convection.

biomass burning outflow. The column mean OH concentrations are $1.4x10^6$, $1.5x10^6$, and $1.6x10^6$ molecules cm^{-3} for the lower, middle, and upper one-third quantiles, respectively, of C_2H_2, corresponding to a relatively small enhancement of 7% due to biomass burning outflow. The small increase in the OH concentration reflects in part the compensating effects of concurrent increases of NO_x and CO concentrations with C_2H_2; NO_x tends to increase OH concentrations, while CO has the opposite effect.

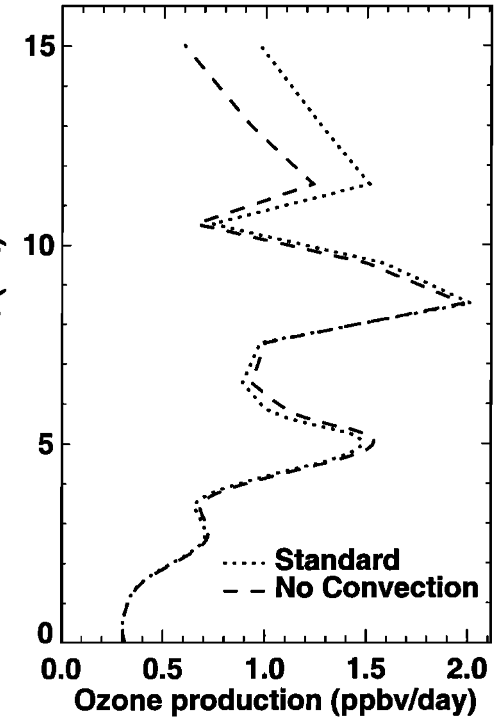


Figure 8. Same as Figure 7 but for the O_3 production rate.

Figure 9 shows simulated production and loss rate and the chemical lifetime of O_3 for the middle one-third quantile of C_2H_2. The production rate of O_3 increases slightly with altitude and is more uniformly distributed vertically than the loss rate. The loss of O_3 occurs primarily in the lowest 5 km. Integrated over the air column of 0-12 km, about 70% of O_3 loss occurs at 0-3 km and 20% occurs at 4-5 km. The column O_3 production (0-12 km) is $1.7x10^{11}$ molecules cm^{-2} s^{-1}, less than half of the column O_3 loss of $4.5x10^{11}$ molecules cm^{-2} s^{-1} (Table 2). About 70% of the O_3 loss is due to the reaction of O^1D+H_2O. The remaining fraction is due to O_3 reactions with OH and HO_2. The chemical lifetime of O_3 increases from 4.5 days near the surface to over 300 days at 15 km.

Air masses with higher concentrations of C_2H_2 have higher concentrations of NO (Figure 2), which is produced most likely from biomass burning and lightning during convection. Higher NO concentrations tend to increase in situ production of O_3. The column O_3 production of 0-12 km are $1.4x10^{11}$, $1.7x10^{11}$, and $3.0x10^{11}$ molecules cm^{-2} s^{-1} for the lower, middle, and upper one-third quantiles, respectively, of C_2H_2 (Table 2). Assuming as previously that the lower one-third quantile of C_2H_2 data represent air masses not significantly influenced by outflow from biomass burning regions, we estimate that the column production of O_3 is increased by 45%, or $6.3x10^{10}$ molecules cm^{-2} s^{-1}, due to biomass burning outflow.

Our estimate of the column O_3 production of $1.7x10^{11}$ molecules cm^{-2} s^{-1} compares well with $2.0x10^{11}$ molecules cm^{-2} s^{-1} over the tropical North Pacific calculated by *Davis et al.* [1996a] and $1.4\text{-}1.8x10^{11}$ molecules cm^{-2} s^{-1} over the tropical South Pacific calculated by *Schultz et al.* [1999]. Our estimated column O_3 loss of $4.5x10^{11}$ molecules cm^{-2} s^{-1} is, however, significantly higher than $2.9x10^{11}$ molecules cm^{-2} s^{-1} calculated by *Davis et al.* [1996a] and $3.2x10^{11}$ molecules cm^{-2} s^{-1} calculated by *Schultz et al.* [1999]. Compared with median values compiled by *Davis et al.* [1996a] (for PEM-West A), median O_3 and H_2O concentrations for the middle one-third quantile of C_2H_2 used in our model during PEM-

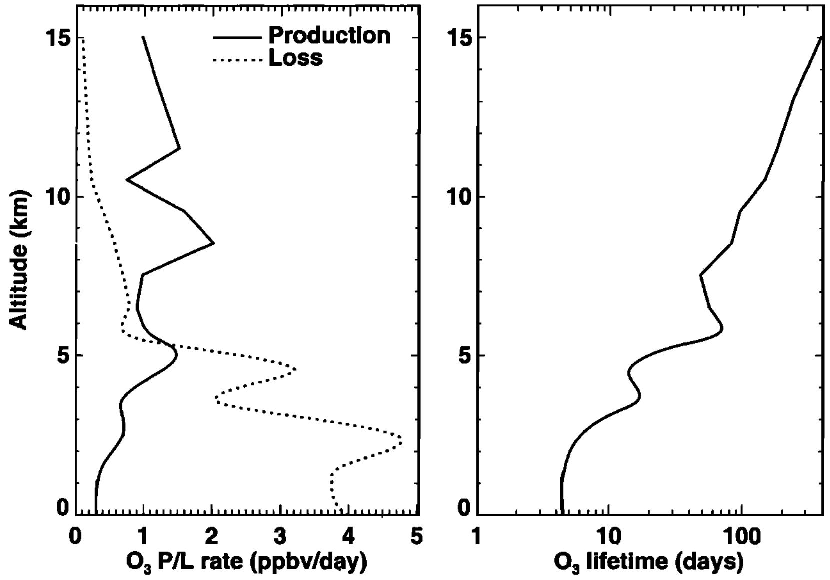


Figure 9. Simulated O_3 production and loss rate and its chemical lifetime as a function of altitude for the middle one-third quantile of C_2H_2.

Tropics A (Figure 2) are both up to a factor of 2 greater at 0-4 km. The difference between our estimate and that of *Schultz et al.* [1999] largely reflects the difference in the data selection method and the large variability in observed H_2O concentrations. Data shown in Figure 9 correspond to median concentrations for the middle one-third quantile of C_2H_2. Median concentrations of O_3 and H_2O selected in this manner tend to be higher than the median concentrations for all the data. The difference is much larger for H_2O than O_3. For instance, median H_2O concentrations are factors of 2 and 6 higher at 2-3 and 4-5 km, respectively, for the middle one-third quantile of C_2H_2 than for all the data. High H_2O concentrations at these altitudes are reflected in the O_3 loss rate in Figure 9. The uncertainty in the model estimate of the column O_3 loss tends to be larger than that of the column O_3 production since the former estimate relies largely on a subset of observations (below 3 km) that also have large variability.

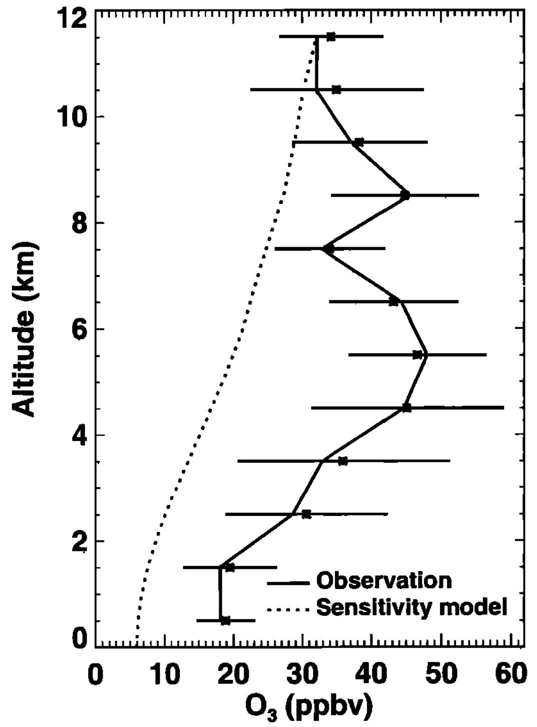


Figure 10. Comparison of observed and simulated vertical profiles O_3 for the middle one-third quantile of C_2H_2. Symbols are the same as in Figure 1. The concentration of O_3 at 11-12 km is specified as observed in the sensitivity simulation.

The net column O_3 deficits (0-12 km) computed in our model are $2.3x10^{11}$, $2.8x10^{11}$, and $2.4x10^{11}$ molecules cm^{-2} s^{-1} for the lower, middle, and upper one-third quantiles, respectively, of C_2H_2. The largest deficit is for the middle one-third quantile of C_2H_2 due to a combination of high relative humidity and relatively high O_3 concentrations (Figures 2 and 4). *Schultz et al.* [1999] suggested that the large O_3 deficit is supplied mostly by longitudinal transport of O_3 into the region. We conduct a sensitivity simulation; the concentration of O_3 at 11-12 km is specified as observed, below which altitude O_3 concentrations are determined by chemical production and loss and a downward flux of O_3 from the upper troposphere due to subsidence. The resulting O_3 concentrations are much lower than the observations (Figure 10). The O_3 column of 0-12 km in the sensitivity simulation is 50% lower, suggesting large amounts of lateral import of O_3 in the region, in agreement with *Schultz et al.* [1999].

5.3. Nitrogen Oxides

Simulated HNO_3 concentrations are generally within the range of observations for the lower and middle one-third quantiles of C_2H_2 but are about a factor of 2 higher above 4 km for the upper one-third quantile of C_2H_2 (Figure 11). Using rate constants for reactions of OH with NO_2 and HNO_3 from *Brown et al.* [1999a, b], instead of those recommended by *Demore et al.* [1997], decreases HNO_3 concentrations by 5-20

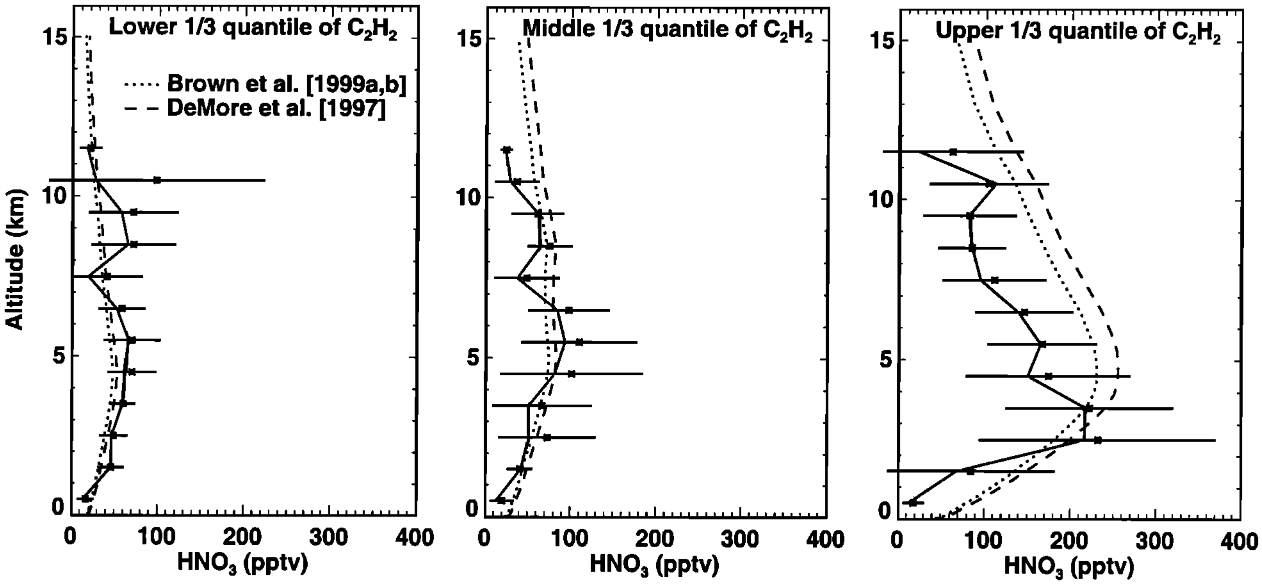


Figure 11. Same as Figure 3 but for HNO_3. Model simulations are shown for using reaction rate constants of OH with NO_2 and HNO_3 from *Brown et al.* [1999a, b] and *DeMore et al.* [1997].

pptv. The relative reduction of HNO_3 concentrations increases with altitude to 20-30% in the upper troposphere. However, only in the case of the upper one-third quantile of C_2H_2 is the agreement between model and observations markedly improved when using the new rate constants.

Model simulations of HNO_3 in the troposphere generally overestimate observations by more than a factor of 2 on a global scale [e.g., *Wang et al.*, 1998b; *Lawrence and Crutzen*, 1998; *Hauglustaine et al.*, 1998]. A number of hypotheses have been proposed to explain this discrepancy [e.g., *Schultz et al.*, 2000]. As in other models, ours does not resolve fractionation of HNO_3 into the aerosol phase. However, *Schultz et al.* [2000] found that aerosol nitrate concentrations were very low during PEM-Tropics A. The hypothesis that rapid chemical recycling of HNO_3 to NO_x in sulfate aerosols or on soot [*Chatfield*, 1994; *Fan et al.*, 1994; *Hauglustaine et al.*, 1996; *Jacob et al.*, 1996; *Lary et al.*, 1997] is compatible with our results because both sulfur and soot are emitted from biomass burning. *Wang et al.* [1998b] found some evidence for the chemical recycling in biomass burning outflow over the tropical South Atlantic during TRACE-A. Loss of HNO_3 due to deposition to cirrus ice crystals followed by subsequent removal by gravitational settling [*Lawrence and Crutzen*, 1998] cannot explain the dependence of HNO_3 overestimates on high C_2H_2 concentrations and appears to be inconsistent with our results. Another explanation for our results could be that air masses with high C_2H_2 concentrations are recently exported from biomass burning regions and have not reached chemical steady state with respect to HNO_3.

Observed peroxyacetylnitrate (PAN) concentrations above 4 km increase from 20 pptv in the lower one-third quantile of C_2H_2 to about 100 pptv in the upper one-third quantile of C_2H_2 (Figure 12). Concentrations of PAN are much lower below 4 km due in part to its short lifetime (20 min to 1 day) against thermolysis. The rate constant of thermolysis decreases exponentially with decreasing temperature and becomes insignificant above 7 km in comparison with photolysis [*Talukdar et al.*, 1995]. The chemical lifetime of PAN in the upper troposphere, controlled by photolysis, is much longer (about 1 month). Model results compare reasonably well with observations for the lower one-third quantile of C_2H_2 but tend to be much lower (up to 40 pptv) for the middle and upper one-third quantile of C_2H_2. The model underestimates likely reflect the influx of relatively long-lived PAN similar to the influxes of O_3 (section 5.2) and C_2H_2 (section 4) associated biomass burning outflow. *Schultz et al.* [1999] suggested that PAN transported from biomass burning regions provides the major source of NO_x in the lower troposphere.

6. Conclusions

We applied a one-dimensional model with an explicit formulation for convective transport to analyze observations over the tropical Pacific during PEM-Tropics A. The vertical transport of the model is constrained by the observations of CH_3I, which originates primarily from the ocean and has a short lifetime of 2-3 days in the tropics. Observed CH_3I concentrations decrease rapidly with altitude in the lowest 3 km but show little dependence on altitude above 5 km. Using diffusive transport alone in the model cannot reproduce the observed profile. When tropical convective mass fluxes from a GCM (GISS model II') are applied, the model grossly overestimates CH_3I concentrations in the upper troposphere but largely underestimates CH_3I in the middle troposphere, indicating that convective mass outflux over the tropical ocean is more evenly distributed with altitude than suggested by the GISS statistics. In order to match observed CH_3I concentrations, the GISS convective outflux above 3 km is redistributed evenly (by air density), yielding a uniform convective turnover lifetime of 20 days. Observations of CH_3I above 12 km, not available during PEM-Tropics A, are necessary to constrain the convective turnover lifetime there.

The model generally reproduces observations of H_2O_2 and CH_3OOH. We find that efficient convective scavenging of

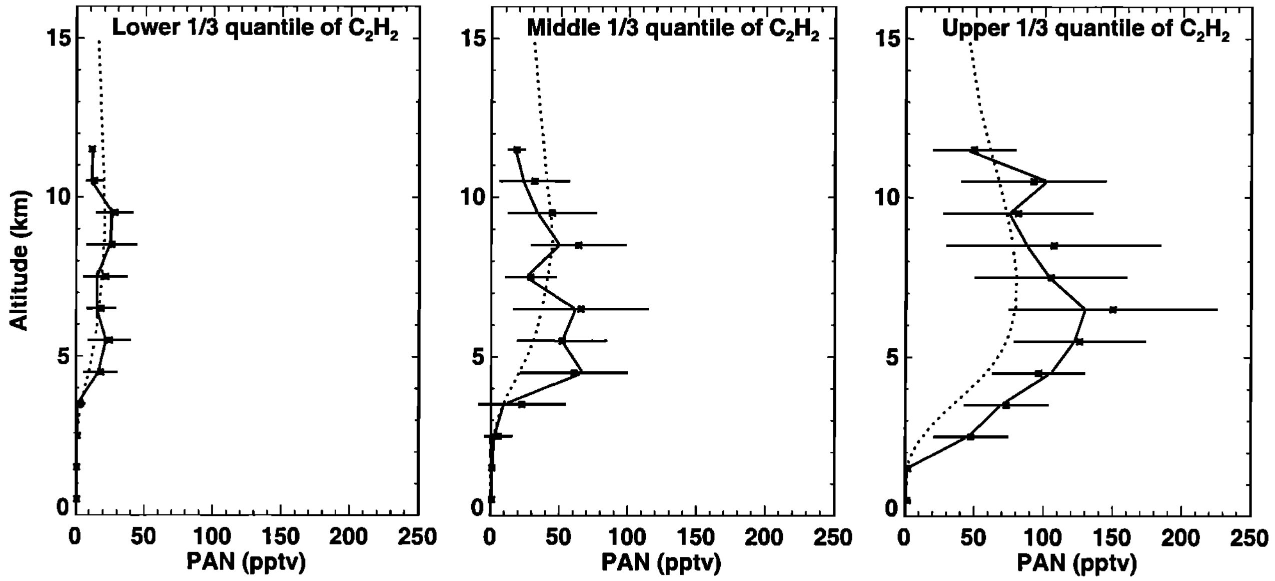


Figure 12. Same as Figure 3 but for peroxyacetylnitrate (PAN). Simulated PAN concentrations are from the standard model.

H_2O_2 by 50-100% is necessary in the model to match observations. Model results indicate that 40-80% of CH_3OOH in the upper troposphere is due to convective transport from the lower troposphere. Photolysis of convected CH_3OOH more than doubles the primary source of HO_x in the region. The reaction of O^1D+H_2O, which is the dominant primary HO_x source in the lower and middle troposphere, is surpassed by photolysis of convected CH_3OOH and photolysis of acetone above 11 km. The latter two sources are also much larger than that from photolysis of convected CH_2O. Convection of CH_3OOH increases the total HO_x source and OH concentration by up to a factor of 3 and 50%, respectively, in the upper troposphere; it increases the production rate of O_3 by about 0.4 ppbv d^{-1} above 11 km. Integrating over the air column of 0-16 km, we find that convection of CH_3OOH has a negligible effect on the column mean OH concentration and increases column O_3 production by only 4%.

In comparison, the effect of biomass burning outflow is much more significant in changing chemistry over the tropical Pacific. We group observations into three equally divided quantiles of C_2H_2 concentrations, because C_2H_2 is a good tracer for biomass burning outflow during PEM-Tropics A. Using the lower one-third quantile of C_2H_2 data as the air masses least influenced by biomass burning outflow, we estimate that 60% (or 9 DU) of O_3 and 75% NO_x enhancements in the air column of 0-12 km are due to import of pollutants from biomass burning regions. The NO_x enhancement increases column O_3 production (0-12 km) by 45%. The import of pollutants increases the column mean OH concentration by a moderate 7% due in part to the offsetting effects of concurrent increases in NO_x and CO concentrations.

The effect of pollutant import into the region is reflected in the budget and concentrations of long-lived chemical species. Whereas short-lived H_2O_2 and CH_3OOH concentrations are generally reproduced based on in situ chemistry and local convective transport within the PEM-Tropics A region, the budget of longer-lived O_3 shows a column O_3 loss a factor of 2 larger than column O_3 production. A model sensitivity study indicates that 50% of O_3 is supplied by lateral import, similar to the finding by *Schultz et al.* [1999]. Peroxyacetylnitrate has a lifetime of about 1 month in the tropical upper troposphere. Model simulations are in much better agreement with PAN observations for the lower one-third quantile of C_2H_2 than those for the middle and upper one-third quantiles of C_2H_2. The underestimate of PAN concentrations for the middle and upper one-third quantiles of C_2H_2 is consistent with pollutant import into the region. Nitric acid also has a long chemical lifetime but is removed efficiently by wet scavenging. As a result, significant direct import of HNO_3 from biomass burning regions is unlikely. Model results are in reasonable agreement with observations for the lower and middle one-third quantiles of C_2H_2. However, the model overestimates HNO_3 concentrations by about a factor of 2 above 4 km for the upper one-third quantile of C_2H_2, as observed previously in other regions of the globe. This discrepancy cannot be explained by recent kinetic data for the reactions of OH with HNO_3 and NO_2 [*Brown et al.*, 1999a, b]. There are two likely explanations: (1) the outflow from biomass burning regions has not reached steady state with respect to HNO_3; and (2) HNO_3 is recycled rapidly back to NO_x by some heterogeneous mechanism in biomass burning outflow.

Acknowledgments. We thank Michael Prather and Martin Schultz for providing useful data. Martin Schultz and anonymous reviewers provided helpful suggestions. We also acknowledge the PEM-Tropics A scientists, DC-8 flight crews, and support staff. In particular, we thank John Bradshaw, Brian Heikes, Glen Sachse, Richard Shetter, Hanwant Singh, and Robert Talbot for their measurements. This work was supported in part by the National Aeronautics and Space Administration (NASA GTE NAG-1-1822).

References

Andreae, M. O., Ocean-atmosphere interactions in the global biogeochemical sulfur cycle, *Mar. Chem.*, *30*, 1-29, 1990.

Andreae, M. O., E. Atlas, H. Cachier, W. R. Cofer III, G. W. Harris,

G. Helas, R. Koppmann, J.-P. Lacaux, and D. W. Ward, Trace gas and aerosol emissions from savanna fires, in *Biomass Burning and Global Change*, vol. 1, *Remote Sensing, Modeling and Inventory Development, and Biomass Burning in Africa*, edited by J. S. Levine, chap. 27, pp. 278-295, MIT Press, Cambridge, Mass., 1996.

Arnold, F., J. Schneider, K. Gollinger, H. Schlager, P. Schulte, E. Hage, P. D. Whitefield, and P. van Velthoven, Observations of upper tropospheric sulfur dioxide- and acetone-pollution: Potential implications for hydroxyl radical and aerosol formation, *Geophys. Res., Lett., 24*, 57-60, 1997.

Atkinson, R., et al., Evaluated kinetic and photochemical data for atmospheric chemistry, *J. Phys. Chem. Ref. Data, 26*, suppl. VI, 1329-1499, 1997.

Bates, T. S., B. K. Lamb, A. Guenther, J. Dignon, and R. E. Stoiber, Sulfur emissions to the atmosphere from natural sources, *J. Atmos. Chem., 14*, 315-337, 1992.

Blake, N. J., D. R. Blake, B. C. Sive, T.-Y. Chen, F. S. Rowland, J. E. Collins, Jr., G. W. Sachse, and B. E. Anderson, Biomass burning emissions and vertical distribution of atmospheric methyl halides and other reduced carbon gases in the South Atlantic region, *J. Geophys. Res., 101*, 24,151-24,164, 1996.

Blake, N. J., et al., Influence of southern hemispheric biomass burning on midtropospheric distributions of nonmethane hydrocarbons and selected halocarbons over the remote South Pacific, *J. Geophys. Res., 104*, 16,213-16,232, 1999.

Brown, S. S., R. K. Talukdar, and A. R. Ravishankara, Rate constants for the reaction $OH + NO_2 + M \rightarrow HNO_3 + M$ under atmospheric conditions, *Chem. Phys. Lett., 299*, 277-284, 1999a.

Brown, S. S., R. K. Talukdar, and A. R. Ravishankara, Reconsideration of the rate constant for the reaction of hydroxyl radicals with nitric acid, *J. Phys. Chem., 103*, 3031-3037 , 1999b.

Chatfield, R. B., Anomalous HNO_3 /NO_x ratio of remote tropospheric air: Conversion of nitric acid to formic acid and NO_x ?, *Geophys. Res. Lett., 21*, 2705-2708, 1994.

Chatfield, R. B., and P. J. Crutzen, Sulfur dioxide in remote oceanic air: Cloud transport of reactive precursors, *J. Geophys. Res., 89*, 7111-7132, 1984.

Cohan, D. S., M. G. Schultz, D. J. Jacob, B. G. Heikes, and D. R. Blake, Convective injection and photochemical decay of peroxides in the upper troposphere: Methyl iodide as a tracer of marine convection, *J. Geophys. Res., 104*, 5717-5724, 1999.

Crawford, J., et al., Assessment of upper tropospheric HO_x sources over the tropical Pacific based on NASA GTE/PEM data: Net effect on HO_x and other photochemical parameters, *J. Geophys. Res., 104*, 16,255-16,274, 1999.

Crutzen, P. J., and M. O. Andreae, Biomass burning in the tropics: Impact on atmospheric chemistry and biogeochemical cycles, *Science, 250*, 16,769-16,778, 1990.

Davis, D. D., et al., Assessment of ozone photochemistry in the western North Pacific as inferred from PEM-West A observations during the fall 1991, *J. Geophys. Res., 101*, 2111-2134, 1996a.

Davis, D. D., J. Crawford, S. Liu, S. McKeen, A. Bandy, D. Thornton, and D. Blake, Potential impact of iodine on tropospheric levels of ozone and other critical oxidants, *J. Geophys. Res., 101*, 2135-2147, 1996b.

DeMore, W. B., et al., Chemical kinetics and photochemical data for use in stratospheric modeling, *JPL Publ. 97-4*, 266 pp., 1997.

Dickerson, R. R., et al., Thunderstorms: An important mechanism in the transport of air pollutants, *Science, 35*, 460-465, 1987.

Fan, S.-M., D. J. Jacob, D. L. Mauzerall, J. D. Bradshaw, S. T. Sandholm, D. R. Blake, H. B. Singh, R. W. Talbot, G. L. Gregory, and G. W. Sachse, Origin of tropospheric NO_x over subarctic eastern Canada in summer, *J. Geophys. Res., 99*, 16,867-16,877, 1994.

Fishman, J., C. E. Watson, J. C. Larsen, and J. A. Logan, Distribution of tropospheric ozone determined from satellite data, *J. Geophys. Res., 95*, 3599-3617, 1990.

Fishman, J., J. M. Hoell, R. D. Bendura, R. J. McNeal, and V. W. J. H. Kirchhoff, NSAS GTE TRACE A Experiment (September-October 1992): Overview, *J. Geophys. Res., 101*, 23,865-23,879, 1996.

Gidel, L. T., Cumulus cloud transport of transient tracers, *J. Geophys. Res., 88*, 6587-6599, 1983.

Gierczak, T., J. B. Burkholder, S. Bauerle, and A. R. Ravishankara, Photochemistry of acetone under tropospheric conditions, *Chem. Phys., 231*, 229-244, 1998.

Happell, J. D., and D. W. R. Wallace, Methyl idode in the Greenland/Norwegian Seas and the tropical Atlantic Ocean: Evidence for photochemical production, *Geophys. Res. Lett., 23*, 2105-2108, 1996.

Hauglustaine, D. A., B. A. Ridley, S. Solomon, P. G. Hess, and S. Madronich, HNO_3 /NO_x ratio in the remote troposphere during MLOPEX 2: Evidence for nitric acid reduction on carbonaceous aerosols, *Geophys. Res. Lett., 23*, 2609-2612, 1996.

Hauglustaine, D. A., G. P. Brasseur, S. Walters, P. J. Rasch, J. F. Müller, L. K. Emmons, and M. A. Carroll, MOZART, a global chemical transport model for ozone and related chemical tracers, 2, Model results and evaluation, *J. Geophys. Res., 103*, 28,291-28,335, 1998.

Hoell, J. M., D. D. Davis, D. J. Jacob, M. O. Rodgers, R. E. Newell, H. E. Fuelberg, R. J. McNeal, J. L. Raper, and R. J. Bendura, The Pacific Exploratory Mission in the tropical Pacific: PEM-Tropics A, August-September 1996, *J. Geophys. Res., 104*, 5567-5583, 1999.

Jacob, D. J., et al., Origin of ozone and NO_x in the tropical troposphere: A photochemical analysis of aircraft observations over the South Atlantic basin, *J. Geophys. Res., 101*, 24,235-24,250, 1996.

Jaeglé, L., et al., Observed OH and HO_2 in the upper troposphere suggest a major source from convective injection of peroxides, *Geophys. Res. Lett., 24*, 3181-3184, 1997.

Jaeglé, L., D. J. Jacob, Y. Wang, A. J. Weinheimer, B. A. Ridley, T. L. Campos, G. W. Sachse, and D. E. Hagen, Sources and chemistry of NO_x in the upper troposphere over the United States, *Geophys. Res. Lett., 25*, 1705-1708, 1998.

Kleinman, L. I., Low and high NO_x tropospheric photochemistry, *J. Geophys. Res., 99*, 16,831-16,838, 1994.

Lary, D. J., A. M. Lee, R. Toumi, M. J. Newchurch, M. Pirre, and J. B. Renard, Carbon aerosols and atmospheric photochemistry, *J. Geophys. Res., 102*, 3671-3682, 1997.

Lawrence, M. G., and P. J. Crutzen, The impact of cloud particle gravitational settling on soluble trace gas distributions, *Tellus Ser. B, 50*, 263-289, 1998.

Lind, J. A., and G. L. Kok, Henry's law determinations for aqueous solutions of hydrogen peroxide, methylhydroperoxide, and peroxyacetic acid, *J. Geophys. Res., 91*, 7889-7895, 1986.

Liu, S. C., J. R. McAfee, and R. J. Cicerone, Radon 222 and tropospheric vertical transport, *J. Geophys. Res., 89*, 7291-7297, 1984.

Logan, J. A., M. J. Prather, S. C. Wofsy, and M. B. McElroy, Tropospheric chemistry: A global perspective, *J. Geophys. Res., 86*, 7210-7254, 1981.

McKeen, S. A., et al., Photochemical modeling of hydroxyl and its relationship to other species during the Tropospheric OH Photochemistry Experiment, *J. Geophys. Res., 102*, 6467-6493, 1997a.

McKeen, S. A., T. Gierczak, J. B. Burkholder, P. O. Wennberg, T. F. Hanisco, E. R. Keim, R.-S. Gao, S. C. Liu, A. R. Ravishankara, and D. W. Fahey, The photochemistry of acetone in the upper troposphere: A source of odd-hydrogen radicals, *Geophys. Res. Lett., 24*, 3177-3180, 1997b.

Müller, J.-F., and G. Brasseur, Sources of upper tropospheric HO_x: A three-dimensional study, *J. Geophys. Res., 104*, 1705-1715, 1999.

Olson, J. R., B. A. Baum, D. R. Cahoon, and J. H. Crawford, Frequency and distribution of forest, savanna, and crop fires over tropical regions during PEM-Tropics A, *J. Geophys. Res., 104*, 5865-5876, 1999.

Pickering, K. E., R. R. Dickerson, G. J. Huffman, J. F. Boatman, and A. Schanot, Trace gas transport in the vicinity of frontal convective clouds, *J. Geophys. Res., 93*, 759-773, 1988.

Pickering, K. E., A. M. Thompson, J. R. Scala, W.-K. Tao, R. R. Dickerson, and J. Simpson, Free tropospheric ozone production following entrainment of urban plumes into deep convection, *J. Geophys. Res., 97*, 17,985-18,000, 1992.

Prather, M. J., and D. J. Jacob, A persistent imbalance in HO_x and NO_x photochemistry of the upper troposphere driven by deep tropical convection, *Geophys. Res. Lett., 24*, 3189-3192, 1997.

Rind, D., and J. Lerner, Use of on-line tracers as a diagnostic tool in general circulation model development, 1, Horizontal and vertical transport in the troposphere, *J. Geophys. Res., 101*, 12,667-12,683, 1996.

Schultz, M., et al., On the origin of tropospheric ozone and NO_x over the tropical South Pacific, *J. Geophys. Res., 104*, 5829-5843, 1999.

Schultz, M., D. J. Jacob, J. D. Bradshaw, S. T. Sandholm, J. E. Dibb,

R. W. Talbot, and H. B. Singh, Chemical NO_x budget in the upper troposphere over the tropical South Pacific, *J. Geophys. Res.*, in press, 2000.

Singh, H. B., M. Kanakidou, P. J. Crutzen, and D. J. Jacob, High concentrations and photochemical fate of oxygenated hydrocarbons in the global troposphere, *Nature, 378,* 50-54, 1995.

Smyth, S., et al., Comparison of free troposphere western Pacific air mass classification schemes for the PEM-West A experiment, *J. Geophys. Res., 101,* 1743-1762, 1996.

Talbot, R. W., J. E. Dibb, E. M. Scheuer, D. R. Blake, N. J. Blake, G. L. Gregory, G. W. Sachse, J. B. Bradshaw, S. T. Sandholm, and H. B. Singh, Influence of biomass combustion emissions on the distribution of acidic trace gases over the southern Pacific basin during austral springtime, *J. Geophys. Res., 104,* 5623-5634, 1999.

Talukdar, R. K., J. B. Burkholder, A. M. Schmoltner, J. M. Roberts, R. R. Wilson, and A. R. Ravishankara, Investigation of the loss processes for peroxyacetyl nitrate in the atmosphere: UV photolysis and reaction with OH, *J. Geophys. Res., 100,* 14,163-14,173, 1995.

Trainer, M., E. Y. Hsie, S. A. McKeen, T. Tallamraju, D. D. Parish, F. C. Fehsenfeld, and S. C. Liu, Impact of natural hydrocarbons on hydroxyl and peroxy radicals at a remote site, *J. Geophys. Res., 92,* 11,879-11,894, 1987.

Trainer, M., et al., Observations and modeling of the reactive nitrogen photochemistry at a rural site, *J. Geophys. Res., 96,* 3045-3063, 1991.

Wang, Y., D. J. Jacob, and J. A. Logan, Global simulation of tropospheric O_3-NO_x-hydrocarbon chemistry, 1, Model formulation, *J. Geophys. Res., 103,* 10,713-10,725, 1998a.

Wang, Y., J. A. Logan, and D. J. Jacob, Global simulation of tropospheric O_3-NO_x-hydrocarbon chemistry, 2, Model evaluation and global ozone budget, *J. Geophys. Res., 103,* 10,727-10,755, 1998b.

Wennberg, P. O., et al., Hydrogen radicals, nitrogen radicals, and the production of ozone in the middle and upper troposphere, *Science, 279,* 49-53, 1998.

D. R. Blake and T.-Y. Chen, Department of Chemistry, University of California, Irvine, CA 92697-2025.

S. C. Liu, S. T. Sandholm, and H. Yu, School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332-0340.

Y. Wang, Department of Environmental Sciences, Rutgers University, 14 College Farm Road, New Brunswick, NJ 08901-8551. (yhw@envsci.rutgers.edu)

(Received July 6, 1999; revised October 5, 1999;
accepted November 11, 1999.)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9335–9348, APRIL 16, 2000

Minor flux adjustment near mixing ratio extremes for simplified yet highly accurate monotonic calculation of tracer advection

Chris J. Walcek

Atmospheric Sciences Research Center, State University of New York at Albany

Abstract. A simplified but very accurate method for calculating advection of mixing ratios in a mass conservative and absolutely monotonic manner in divergent or nondivergent multidimensional flows is presented. This scheme uses a second-order-accurate, upstream approximation with monotone limiters and additionally adjusts fluxes at two cell edges around local extremes of a tracer distribution to significantly improve overall advection calculations. The minor flux adjustment slightly aggregates mass around local peaks in a manner which counters the inherent numerical diffusion associated with most numerical advection algorithms when advecting poorly resolved features. When advecting tracer shapes which are resolved by fewer than 10-20 grid cells, this scheme is significantly more accurate than higher-order algorithms for a wide range of test problems. For well-resolved tracer distributions this algorithm is very accurate and usually preserves local peak and minimum values almost perfectly. The scheme is positive-definite, but negative values can be advected with no modifications. A generalized algorithm and FORTRAN subroutine is presented for advecting mixing ratios or other conservative quantities through variable-spaced grids of one to three dimensions, including deformational flows. One- and two-dimensional tests are presented and compared with other higher-order algorithms. The computational requirements of this algorithm are significantly lower than those of other higher-order and less accurate schemes.

1. Introduction

For many geophysical applications it is necessary to calculate the transport of trace constituent concentrations within a moving fluid. The time (t) tendency of a tracer concentration C in a fluid at velocity u in one direction x is

$$\frac{\partial C}{\partial t} = -\frac{\partial (uC)}{\partial x} \quad . \tag{1}$$

For many applications, velocity and concentration distributions are complex and changing in time, requiring numerical approximations of (1) to quantify advection. Unfortunately, many numerical approximations are extremely inaccurate, with solutions containing appreciable numerical diffusion, oscillatory behavior, and physically unreasonable (e.g., negative) concentrations, especially near regions of strong tracer gradients. Ample reviews of numerical advection methods [e.g., *Rood*, 1987] have noted that tracer mass conservation, sign preservation, monotonicity (peak and minima preservation with no spurious peaks calculated), low numerical diffusion, and algorithm simplicity are all desirable features of advection algorithms. Unfortunately, many algorithms used in geophysical models contain one or more of the above listed shortcomings to some degree.

Various numerical solutions to (1) have been developed and improved over the years [*Smolarkiewicz*, 1983; *Colella and Woodward*, 1984; *Tremback et al.*, 1987; *Bott* 1989a, b, 1992; *Carpenter et al.*, 1990]. Some of these flux-correcting schemes assume a parabolic or higher-order polynomial distribution of tracer within cells, and sometimes flux limitations are imposed to ensure a positive-definite solution or to reduce undesirable nonmonotonic errors. Recent versions of these algorithms contain relatively little numerical diffusion [e.g., *Smolarkiewicz and Grabowski*, 1990; *Bott*, 1993]. *Dabdub and Seinfeld* [1994] and *Chock* [1991] suggest that the most accurate advection algorithms are extremely high order schemes [*Gasdag*, 1973] combined with sophisticated filtering procedures. *Prather* [1986] describes a multimoment scheme where tracer mass and several spatial derivatives of the tracer distribution are simultaneously advected. However, these algorithms require high memory and calculation resources. *Costa and Sampaio* [1997] derive extremely high order advection algorithms that they claim should be used when advecting poorly resolved features. Unfortunately, all of these high order schemes perform poorly when sharp gradients or small features are advected, producing physically unreasonable and unrealistic nonmonotonic behavior in divergent multidimensional flows, unless ad hoc filtering or other monotonic adjustments are employed. As noted by *Lin et al.* [1994], empirical flux limitation techniques that remove physically unrealistic oscillations usually introduce appreciable numerical diffusion when sharp gradients or poorly resolved features are advected.

A simple method for solving the problem of accurately advecting small features is to increase the number of grid cells of a numerical model so that advected features are better resolved. Unfortunately, decreasing model grid size greatly increases computational expense, and for many pollution and atmospheric chemistry applications, reactive constituents are emitted into flows at individual "points," and no matter how feasibly small the grid size is devised to study the urban, regional

Copyright 2000 by the American Geophysical Union

Paper number 1999JD901142.
0148-0227/00/1999JD901142$09.00

or global impacts of pollutants, the individual "plumes" or "puffs" remain essentially one grid cell in size. Even for many nonchemical applications, such as advection of heat or water vapor in meteorological models, small-scale features are inevitably generated from smooth initial conditions in realistic atmospheric flows through a number of physical processes. These small features or sharp gradients will induce undesirable oscillatory or diffusive behavior during advection calculations, even when the highest-order accurate advection algorithms are used.

Here a simple, absolutely monotonic numerical advection algorithm with extremely low numerical diffusion is presented and tested. The method represents a simplification of the *Walcek and Aleksic* [1998] scheme. The key innovation of both the Walcek and Aleksic scheme and this technique is the imposition of a minor flux adjustment at two cell edges surrounding each local extreme of a tracer distribution. This minor adjustment efficiently limits the initial artificial numerical diffusion away from peaks, appreciably slowing the subsequent cascade of numerical diffusion that spreads mass away from local peaks of a tracer distribution.

2. Algorithm Description

Equation (1) can be expressed in terms of tracer mixing ratio Q (=C/ρ):

$$\frac{\partial(\rho Q)}{\partial t} = -\frac{\partial(u\rho Q)}{\partial x} , \qquad (2)$$

where ρ is the fluid density. Equation (2) is numerically approximated using a forward time difference, and evaluating the space derivative at time t. Mixing ratios at time $t+\Delta t$ can be initially estimated using

$$Q_i^{\text{guess}} = \frac{\left[Q_i^t D_{d-1} - \frac{(\Delta t u \rho Q_f)_{i+1/2}}{\Delta x_i} + \frac{(\Delta t u \rho Q_f)_{i-1/2}}{\Delta x_i} \right]}{D_d} . \qquad (3)$$

Here Q_f is the average mixing ratio within the fluid flowing across each cell edge during time step Δt, and $D_{d\text{-}1}$ and D_d are the dimensionally dependent fluid densities at the beginning and end of the time step or dimensional step in a multidimensional calculation performed one dimension at a time. As shown schematically in Figure 1, the subscript i+1/2 refers to the "higher-i" face of each cell (the edge of cell i nearest to the i+1 cell), and i-1/2 refers to the "lower-i" side of the cell. Two- or three-dimensional advection calculations are performed by applying this one dimensional algorithm sequentially over each dimension. However, as described below, the fluid densities D_d must be sequentially changed for the second and third dimensions (subscript d) even in incompressible fluids, while the ρ terms used to calculate fluxes are held fixed at the initial fluid density for all dimensions. For 1-D advection with constant winds, $D_{d\text{-}1} = D_d = 1$.

Fluxes $(\Delta t u \rho Q_f)_{i\pm 1/2}$ are evaluated using the 2nd-order accurate *van Leer* [1977] approach, which assumes a linear distribution of tracer within each grid cell, and the gradient within the cell, $\partial Q/\partial x = (Q_{i+1} - Q_{i-1})/2\Delta x_i$. This approach is identical to the simplest L=1 area preserving *Bott* [1989b] polynomial, averaged according to *Easter* [1993]. The average mixing ratio Q_f in the fluid transported out of each cell during Δt, shown schematically in Figure 1, is

$$(Q_f)_{i+1/2} = Q_i + \frac{(Q_{i+1} - Q_{i-1})(1-c)}{4}\alpha \qquad u_{i+1/2} \geq 0 \quad (4a)$$

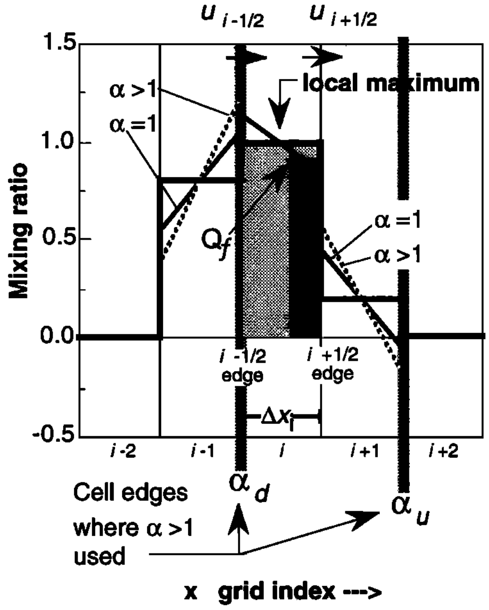


Figure 1. Mixing ratio distribution within grid cells near a local peak. Standard *van Leer* [1977] scheme (solid line, α=1) and "sharpened" (α specified using equation (10) for Courant number 0.4, stippled line) shown. Using α>1 in two cells sandwiching a local maximum greatly reduces numerical diffusion by numerically "aggregating" mass around the local peak before updating mixing ratios, opposing the tendency of numerical diffusion to erroneously spread mass away from local peaks.

$$(Q_f)_{i-1/2} = Q_i + \frac{(Q_{i-1} - Q_{i+1})(1-c)}{4}\alpha \qquad u_{i-1/2} < 0 \quad (4b)$$

where c is the local "upwind" Courant number ($c = u_{i+1/2}\Delta t/\Delta x_i$ if $u_{i+1/2} > 0$ and $c = |u_{i-1/2}|\Delta t/\Delta x_i$ if $u_{i-1/2} < 0$). Setting the "slope factor" α=1 everywhere yields the van Leer or Bott L=1c algorithm. For this algorithm, α=1 everywhere except for two cells sandwiching each local peak or minimum of a tracer distribution (equations (10a) and (10b)). In the unusual case where there is only outflow from a cell ($u_{i+1/2} > 0$ and $u_{i-1/2} < 0$), mixing ratios cannot change during the time step Δt, and $Q_f = Q_i$ at both the lower and higher cell edges. Equation (4) is the fundamental difference between this algorithm and the somewhat more complicated "bilinear" algorithm by *Walcek and Aleksic* [1998]. In that algorithm, two line segments which intersect at the cell center are assumed to describe the tracer distribution within a given cell. It was subsequently discovered that using only a single line segment produced comparable or greater accuracy for all tests considered.

Dimensionally dependent densities (D_d in equation 3) are specified as

$$D_0 = \rho_i \,.$$
$$D_1 = D_0 - [(\rho u)_{i+1/2} - (\rho u)_{i-1/2}]\Delta t/\Delta x_i \,,$$
$$D_2 = D_1 - [(\rho v)_{j+1/2} - (\rho v)_{j-1/2}]\Delta t/\Delta y_j \,,$$
$$D_3 = D_2 - [(\rho w)_{k+1/2} - (\rho w)_{k-1/2}]\Delta t/\Delta z_k \,, \qquad (5)$$

where ρ_i is the initial fluid density, and $\rho_{i\pm 1/2}$ is the density of fluid advected across the specified cell interface. Cross-flowing densities $\rho_{i\pm 1/2}$ can be defined using an upstream approach or other approximations but must be identical to the densities used in defining the tracer fluxes $(\Delta t u \rho Q_f)_{i\pm 1/2}$ in equation (3).

Note that ρ terms in (3) and (5) are the same for each dimension of a multidimensional calculation, while D_d terms change with each dimensional calculation. For three-dimensional calculations, mixing ratios can be first updated doing $x(_i)$ direction advection using u velocities. These updated mixing ratios are then advected in the $y(_j)$ direction using v velocities, followed by $z(_k)$ dimension advection using w velocities. During x advection, D_0 and D_1 are used in (3) for $D_{d\text{-}1}$ and D_d. During y advection with velocities v, D_1 and D_2 are used, and during z advection D_2 and D_3 are used in (3). Using this "operator splitting" approach for multidimensional advection, fluid density can artificially change when performing advection in one dimension, even in nondivergent flows, but as formulated here, any artificial density perturbations are intermediate and are naturally compensated during subsequent dimension calculations. If advection calculations are performed in a mass- or pressure-based (sigma) coordinate system, the ρ terms in (3) and (5) are equivalent to "P-star" (P^*), the pressure difference between the top and base of the model domain.

2.1. Monotonic Constraints

A monotonic advection algorithm will never generate mixing ratios higher than the highest mixing ratio initially present in the mixing ratio field, nor will a mixing ratio ever become lower than the lowest initial mixing ratio. Some methods of enforcing monotonicity consider monotonicity from a total domain perspective [*Lin et al.*, 1994], which can allow local unrealistic oscillatory behavior. Total variational diminishing (TVD) limitations assume that averaged over a domain, gradients between individual cells must never increase, which *Thuburn* [1997] showed was consistent with a general "universal" monotonic limiter. Here monotonic constraints are employed locally on the basis of obvious physical limitations. If the Courant number is <1, at time $t+\Delta t$ it is physically impossible for updated mixing ratios to be greater than the highest mixing ratio of the upwind cell or the mixing ratio of the cell initially. Likewise, updated mixing ratios cannot be lower than the lowest mixing ratio of the upwind cell or the mixing ratio in the cell initially. Monotonicity is guaranteed through two steps. First, Q_f in (4) is constrained to fall in the range of the mixing ratios on either side of the interface where fluxes are evaluated:

$$\min(Q_i, Q_{i+1}) < (Q_f)_{i+1/2} < \max(Q_i, Q_{i+1}) \quad u_{i+1/2} \geq 0, \tag{6a}$$

$$\min(Q_i, Q_{i\text{-}1}) < (Q_f)_{i\text{-}1/2} < \max(Q_i, Q_{i\text{-}1}) \quad u_{i\text{-}1/2} < 0. \tag{6b}$$

Second, outflowing fluxes are adjusted so that updated mixing ratios never exceed or fall below the physically constrained highest ($Q^{t+\Delta t}_{\max}$) and lowest ($Q^{t+\Delta t}_{\min}$) mixing ratios allowed. These upper and lower limits are given by

$$Q^{t+\Delta t}_{\min} = \mathrm{Min}(Q_i, Q_{i\text{-}1}) \quad Q^{t+\Delta t}_{\max} = \mathrm{Max}(Q_i, Q_{i\text{-}1}) \quad u_{i\text{-}1/2} \geq 0, \tag{7a}$$

$$Q^{t+\Delta t}_{\min} = \mathrm{Min}(Q_i, Q_{i+1}) \quad Q^{t+\Delta t}_{\max} = \mathrm{Max}(Q_i, Q_{i+1}) \quad u_{i+1/2} < 0. \tag{7b}$$

At inflow-only cells where $u_{i\text{-}1/2} \geq 0$ and $u_{i+1/2} < 0$, the mixing ratio limits are the maximum or minimum of the three cells $Q_{i\text{-}1}$, Q_i, Q_{i+1} contributing mass to the cell during Δt. If Q^{guess} calculated using (3) violates these monotonic constraints, the updated mixing ratio is set to the violated limit,

$$Q_i^{t+\Delta t} = \mathrm{Max}\left[\mathrm{Min}(Q^{t+\Delta t}_{\max}, Q_i^{\text{guess}}), Q^{t+\Delta t}_{\min}\right] \; ; \tag{8}$$

then the outflowing flux is adjusted so that the $Q^{t+\Delta t}$ equals the violated limit:

$$\begin{aligned}(\Delta t u \rho Q_f)_{i+1/2} &= \left[Q_i^t D_{d-1} - Q_i^{t+\Delta t} D_d\right]\Delta x_i \\ &\quad + (\Delta t u \rho Q_f)_{i-1/2}\end{aligned} \qquad u_{i+1/2} \geq 0, \tag{9a}$$

$$\begin{aligned}(\Delta t u \rho Q_f)_{i-1/2} &= \left[Q_i^{t+\Delta t} D_d - Q_i^t D_{d-1}\right]\Delta x_i \\ &\quad + (\Delta t u \rho Q_f)_{i+1/2}\end{aligned} \qquad u_{i\text{-}1/2} < 0. \tag{9b}$$

2.2. Limiting Numerical Diffusion Near Extremes

Overall numerical diffusion is significantly removed by a minor adjustment of the factor α in (4) near local extremes. If there is a local extreme at the cell immediately downwind of the edge where fluxes are calculated, it was found that using

$$\alpha_d = 1.75 - 0.45c \tag{10a}$$

greatly reduces overall numerical diffusion. This adjustment must be used together with

$$\alpha_u = \max(1.5,\ 1.2 + 0.6c) \ , \tag{10b}$$

if the cell two cells upwind of the face where fluxes are being calculated is a local extreme. In (10), c is the "upwind" Courant number ($c = u_{i\pm1/2}\Delta t/\Delta x_i$ if $u_{i\pm1/2} > 0$ and $c = |u_{i\pm1/2}|\Delta t/\Delta x_i$ if $u_{i\pm1/2} < 0$). Cell edges where $\alpha > 1$ is used near a local peak are shown in Figure 1 for positive winds. As shown in Figure 1, these two adjustments together "sandwich" each local extreme and therefore effectively "aggregate" tracer around local maximums in a manner which counters the numerical diffusion that diffuses mass away from extremes. A local extreme is defined to exist at cell i if $Q_i \geq \max(Q_{i+1}, Q_{i\text{-}1})$ or if $Q_i \leq \min(Q_{i+1}, Q_{i\text{-}1})$. The method for deriving equations (10a) and (10b) is described in the appendix.

2.3. Calculation Sequence

Updated mixing ratios are calculated in a manner similar to that of *Bott* [1992]. An inflow flux is first defined at the lowest cell in a grid, where it is convenient to assume an upstream flux at the domain boundaries. If $u_{1\text{-}1/2}$ is positive, $(Q_f)_{1\text{-}1/2}$ at the lower edge of cell 1 is set to the boundary condition mixing ratio provided as input to the advection calculation. If periodic boundary conditions are assumed, $(Q_f)_{1\text{-}1/2} = Q_{i\max}$, where $Q_{i\max}$ is the mixing ratio at the highest-i cell. Q_f at the outflowing 1+1/2 edge is calculated using (4a), which requires using the appropriate α. In (4), $\alpha = 1$ unless fluxes are being evaluated near an extreme, where (10) is used to define α. Q_f is then monotonically constrained (equation (6a)), then updated mixing ratios for cell 1 are estimated using (3). These updated mixing ratios are monotonically constrained (equation (8)), and if the initial guess violates a monotonic limit, the outflowing flux is adjusted using (9a) to force a monotonic solution. This outflowing flux becomes the inflow flux for updating the mixing ratio at the next higher cell in the grid, and the calculation cycle is repeated up through the grid where winds are positive.

If winds are negative, the calculation cycle is started at the highest cell in the grid, and $(Q_f)_{i\max+1/2}$ is set to the boundary mixing ratio (inflow/outflow boundary conditions) or to Q_1 (periodic boundary conditions). The outflowing mixing ratio at the $i$max-1/2 edge is then calculated using (4b) and monotonically constrained (equation (6b)), then updated mixing ratios for cell $i$max are calculated using (3). These updated mixing ratios are monotonically constrained (8), and if necessary, outflowing fluxes are adjusted using (9b). This outflowing flux becomes the inflow flux for updating mixing ratios at the next lower cell in the grid, and the calculation cycle is repeated for successive lower cells in the grid wherever winds are negative.

For many applications, wind directions change sign within any given 1-D advection line. For example, one may encounter a grid line where the lowest cell wind speed is negative ($u_{1-1/2}<0$) and the highest cell wind speed is positive ($u_{i\max+1/2}>0$). Under these conditions, calculation must begin at a grid where there is only outflow ($u_{i-1/2}<0$ and $u_{i+1/2}>0$). Mixing ratios at this cell will not change during Δt, and fluxes out of this cell are defined using the upstream approach ($Q_f = Q_i$). Fluxes and updated mixing ratios are then calculated and adjusted by marching up from the outflow cell to higher-i cells wherever winds are positive, followed by updating mixing ratios and limiting fluxes going down through the grid from the outflow cell wherever $u<0$. General application of this algorithm to arbitrary flows requires two passes through a grid line: a pass updating mixing ratios proceeding up through the grid where winds are positive, followed by a pass back through the grid updating mixing ratios where winds are negative.

For applications of this algorithm with periodic boundary conditions, one can assume an upstream flux for calculating fluxes at the periodic boundary. However, the upstream flux between cell 1 and cell "$i$max" cannot be adjusted using (9) when the calculation cycle reaches the starting point again at "$i$max." This limitation has the effect of introducing a minor amount of upstream numerical diffusion at one cell interface as features traverse the boundary. In a domain with periodic conditions, calculation can begin at any arbitrary cell, and the outflowing flux at that cell must be specified as upstream. Ideally, one can choose an outflow-only cell for beginning the calculation sequence. If this is not feasible to implement, then fluxes at the 1-1/2 (the first flux calculated) can be calculated as described above by substituting the appropriate periodic indices in place of the i-1, i, and i+1 cells in (4) and then proceeding as described above. However, if the flux at the end of the calculation sequence at the highest cell edge does not equal the initial flux estimate (owing to monotonic adjustment using (9)), then the mixing ratio at the first cell must be recalculated using the adjusted "$i$max+1/2" flux and the already calculated flux at "1+1/2."

3. Results

A wide variety of advection tests and methods for measuring algorithm accuracy have been applied to numerical advection algorithms. This scheme has been subjected to many of the tests of passive tracer transport described by *Smolarkiewicz and Grabowski* [1990], *Bott* [1992], *Rasch* [1994], *Lin and Rood* [1996], *Lin et al.* [1994], *Chock* [1991], *Pietrzak* [1998], *Carpenter et al.* [1990], *Russell and Lerner* [1981], *Thuburn* [1996], *Clappier* [1998], *Dabdub and Seinfeld* [1994], *Takacs* [1985], *Chlond* [1994], *Prather* [1986], *Schneider* [1984] *Costa and Sampaio* [1997], and *Yamartino* [1993]. These tests include various divergent and nondivergent one- and two-dimensional flows, and they also consider several sizes and shapes of tracer distributions. Some tests consider flow on a spherical surface and flows in spherical coordinates, and some consider flows where fluid density varies in time or space. For all of these tests and testing methods the scheme described here produces errors that are comparable to or appreciably smaller than the best results published most of the time. The only schemes which yield better performance than this scheme for some conditions are nonmonotonic schemes which can under other conditions produce unrealistic advection calculations for tracer distributions containing small features or steep gradients, as shown in section 3.4.

Here this scheme is directly compared with a more complicated earlier version of this algorithm [*Walcek and Aleksic*, 1998], referred to in subsequent figures and tables as the "bilinear" scheme. This scheme is also compared with monotonic and nonmonotonic versions of the schemes by *Smolarkiewicz* [1983], *Bott* [1992], an implicit Galerkin finite element technique [*Toon et al.*, 1988], and the multimoment *Prather* [1986] scheme. All tests of the Smolarkiewicz algorithm use three corrective steps, and corrective antidiffusion velocities are multiplied by 1.02 to improve performance relative to no adjustment. Four versions of the successively higher order *Bott* [1989b] scheme are used: L=1, L=2, L=4, and L=8. Coefficients for the Bott L=8 polynomials are presented in *Costa and Sampaio* [1997].

3.1. Error Measures

This algorithm is subjected here to several one- and two-dimensional tests. Six error measures (in percent) are used to quantify algorithm performance:

$$pker = 100\left(\frac{E_{\max} - C_{\max}}{E_{\max} - E_{\min}}\right), \qquad (11)$$

$$mser = 100\left(1 - \sum_i C_i \Big/ \sum_i E_i\right), \qquad (12)$$

$$rmser = 100\,\frac{\sqrt{\sum_i (C_i - E_i)^2/N}}{(E_{\max} - E_{\min})}, \qquad (13)$$

$$dsper = 100\left(1 - \sum_i C_i^2 \Big/ \sum_i E_i^2\right), \qquad (14)$$

$$m\Delta er = 100\left(\frac{\underset{i}{\mathrm{Max}}|C_i - E_i|}{E_{\max} - E_{\min}}\right), \qquad (15)$$

$$mner = 100\left(\frac{C_{\min} - E_{\min}}{E_{\max} - E_{\min}}\right). \qquad (16)$$

These error measures *pker*, *mser*, *rmser*, *dsper*, *mΔer* and *mner* quantify the errors in peak preservation, tracer mass, root-mean-square, dispersion, maximum absolute difference, and minimum preservation respectively. Here C and E are the calculated and exact mixing ratios, and N is the number of cells over which errors are calculated. The max and min subscripts refer to the highest and lowest values in the limited subdomain over which error scores are computed. Most error measures are normalized to the range of the initial tracer distribution and are thus invariant to the background and peak values chosen for the comparison. All sums and error measures are evaluated only over a limited subdomain of the entire domain, and thus the error measures are not a function of the entire test domain size. While all schemes tested here conserve mass absolutely over the entire domain, it is not unusual to see mass displaced away from its proper location through numerical diffusion or

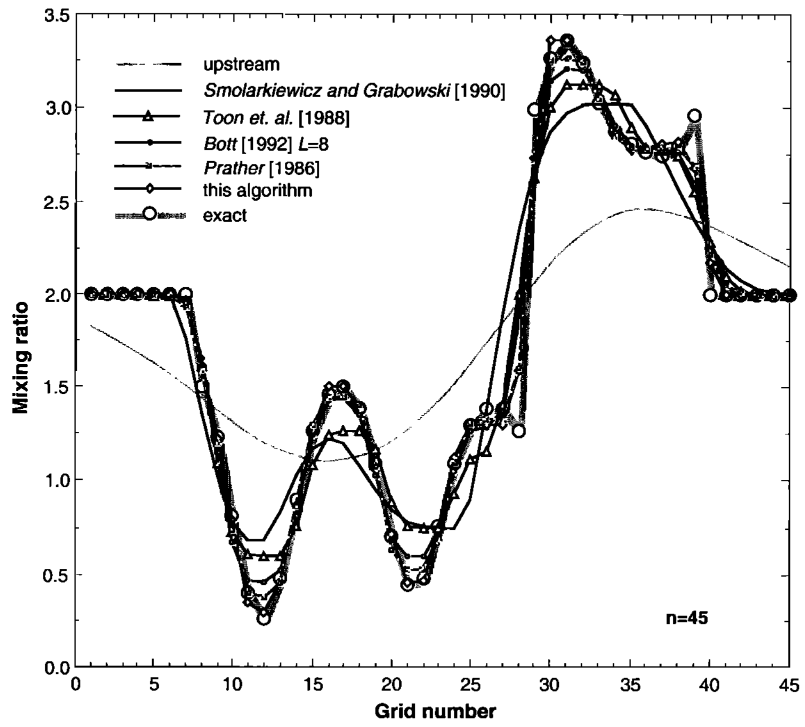


Figure 2. Calculated and exact distribution of mixing ratio advected 50 grid cells at a Courant number 0.2 (250 time steps). This scheme is compared with several other monotonic advection algorithms.

nonmonotonic oscillations. All of these error measures are defined so that a zero error is the best possible score, and the peak and minimum errors yield a negative score if calculated distributions are nonmonotonic. A positive mass error means that mass has diffused away from its proper location, and there is less mass than there should be in the evaluation subdomain.

3.2. One Dimensional Test

Figure 2 shows a fairly rigorous test of one-dimensional advection in constant winds. Here the irregular signal of *Smolarkiewicz and Grabowski* [1990] is advected 50 cells at a Courant number 0.2, requiring 250 time steps. This test uses a domain of 100 cells where constant background mixing ratios are assumed as boundary conditions, and error scores are only computed over the 45-cell subdomain shown in Figure 2.

Error scores for several algorithms subjected to this 1-D test are summarized in Table 1. The rows of Table 1 are sorted by increasing rms errors, but monotonic and nonmonotonic algorithms are sorted separately. The error scores listed in Table 1 show that this algorithm yields errors that are nearly identical

Table 1. Errors Advecting Irregular Signal a Distance of 50 Cells at Courant Number 0.2 and 250 Iterations

Algorithm	*mser*, %	*rmser*, %	*pker*, %	*mner*, %	*dsper*, %	*mΔer*, %
			Monotonic Algorithms			
This algorithm	0.000	2.80	0.00	0.86	0.22	10.40
Prather [1986]	-0.099	2.94	3.15	3.91	0.63	11.14
Walcek and Aleksic [1998]	0.000	3.01	0.00	1.61	0.20	11.08
Bott [1992] *L*=8	0.000	4.62	4.93	6.30	1.55	19.87
Bott [1992] *L*=4	0.000	5.77	6.08	7.40	2.10	24.97
Toon et al. [1988]	0.004	6.50	7.58	10.81	2.23	23.64
Smolarkiewicz [1983]*	0.005	9.20	11.50	12.81	3.01	36.02
Upstream	0.152	17.04	28.97	27.32	13.15	34.98
			Nonmonotonic Algorithms			
Prather [1986]	-0.029	2.24	-0.96	0.19	0.24	7.72
Bott [1989] *L*=8	-0.016	3.91	-2.19	0.93	0.46	15.87
Toon et al.,[1988]	0.143	6.58	0.22	-1.37	0.56	27.48
Smolarkiewicz and Grabowski [1990]*	-0.213	11.12	9.73	9.08	1.20	37.71

*three corrective steps.

to or slightly better than the somewhat more complicated "bilinear" version of this algorithm. Error scores for both old and new/simplified versions of this algorithm are significantly lower, sometimes by factors of 2-5 or more, than the highest-order *Bott* [1992] L=8 scheme or the *Prather* [1986] algorithm. The results of this algorithm shown in Figure 2 are noticeably more accurate than the piecewise parabolic method [PPM, *Colella and Woodward*, 1984] as subjected to this same test by *Lin and Rood* [1996].

While not shown in Figure 2, scores of nonmonotonic versions of the Bott, Prather, Smolarkiewicz and Toon techniques are shown in Table 1 for comparison. The nonmonotonic Prather scheme appears to yield the best scores for this particular test, although not all error measures are superior. The mass error scores for some of the nonmonotonic algorithms demonstrate a sometimes serious deficiency of using nonmonotonic schemes for advection calculations. While nonmonotonic schemes can yield better scores for some error measures relative to the corresponding monotonic versions, nonmonotonic schemes do not conserve mass locally as mass disperses in the ripples and oscillations that can spread appreciable distances from the proper locations of the advected shapes. For the higher-order monotonic schemes tested on this problem, Table 1 shows that at worst, numerical diffusion moves a few parts in 10^4 - 10^5 outside six to seven grid cells of the original tracer distribution initially encompassed by 33 grid cells. In contrast, the nonmonotonic schemes contain mass errors of up to 1-2 parts per thousand, implying that there are minor compensating mass errors outside the subdomain shown in Figure 2. Therefore, while nonmonotonic schemes can achieve excellent scores for this test, these schemes do not exactly conserve mass locally. Nonmonotonic schemes also produce physically absurd and unrealistic local oscillations, which for some tracer distributions may be undesirable, as shown in section 3.4.

The scheme has also been tested for numerous additional Courant numbers and 1-D shapes shown by *Walcek and Aleksic* [1998]. These additional tests consider step, triangular, cosine, Gaussian, exponential, and parabolic shapes of various sizes, and the results are as good as or better than the more complicated "bilinear" version of this algorithm [*Walcek and Aleksic*, 1998], which itself is noticeably more accurate than many higher-order and computationally more expensive schemes for a wide range of test shapes and sizes.

3.3. Two-Dimensional Rotating Plane Tests

This algorithm is subjected to a slightly more rigorous version of the standard rotating plane test considered by others. On a 100Δx by 100Δy plane, five sizes of a particular tracer shape are rotated around the plane six times, using 628 time steps per rotation. The x direction velocity u=-(j-50.5)/50, and the y direction velocity v= (i-50.5)/50, where i and j refer to grid index numbered from the lower left of the counterclockwise rotating plane. The maximum Courant number in any one dimension occurs at the outer edges of the domain and is slightly less than 0.5. In these tests a background mixing ratio of 10 is used, and the initial peak mixing ratio is 100. Figures 3-5 show the rotation of several sizes of cone-shaped (Figure 3), cylinder (Figure 4), and cosine hill (Figure 5) tracer distributions. Moving counterclockwise from the largest shape, the radius of each subsequent shape is reduced by half. Thus five identical shapes of radius 16Δx, 8Δx, 4Δx, 2Δx, and 1Δx are simultaneously displayed in Figures 3-5. All advected shapes are initially located ~30Δx from the center of the plane, and the center or peak value of each shape is one point at (i,j)= (71,29), (60,78), (29,71), (21,43) and (35,24). Figures 3-5 compare this algorithm with a monotonic version of the eighth-order area-preserving polynomial approach of *Costa and Sampaio* [1997].

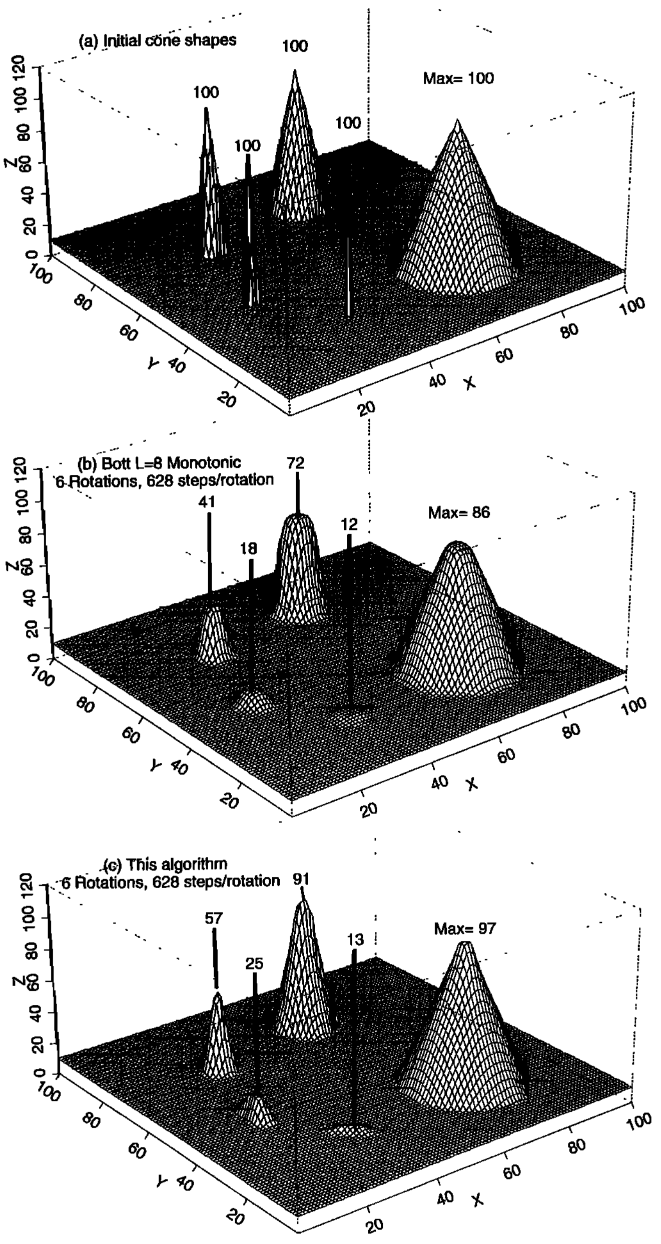


Figure 3. (a) Initial and exact cone-shaped tracer distribution for rotating plane test. (b) Distributions calculated using *Bott* [1992] L=8. (c) Distributions calculated using this algorithm. Calculations are for a distribution of five cones each ~30 cells from domain center. Calculated shapes rotated six times around the domain center using 628 time steps per rotation (3768 time steps). Going counter-clockwise from the largest shape, each smaller shape has a radius one-half the next largest shape of 16Δx, 8Δx, 4Δx, 2Δx and 1Δx in size.

Figure 3 shows that the peak and shape of all cone sizes are advected more accurately than the *Bott* [1992] L=8 approach, especially for the poorly resolved cones. The somewhat more

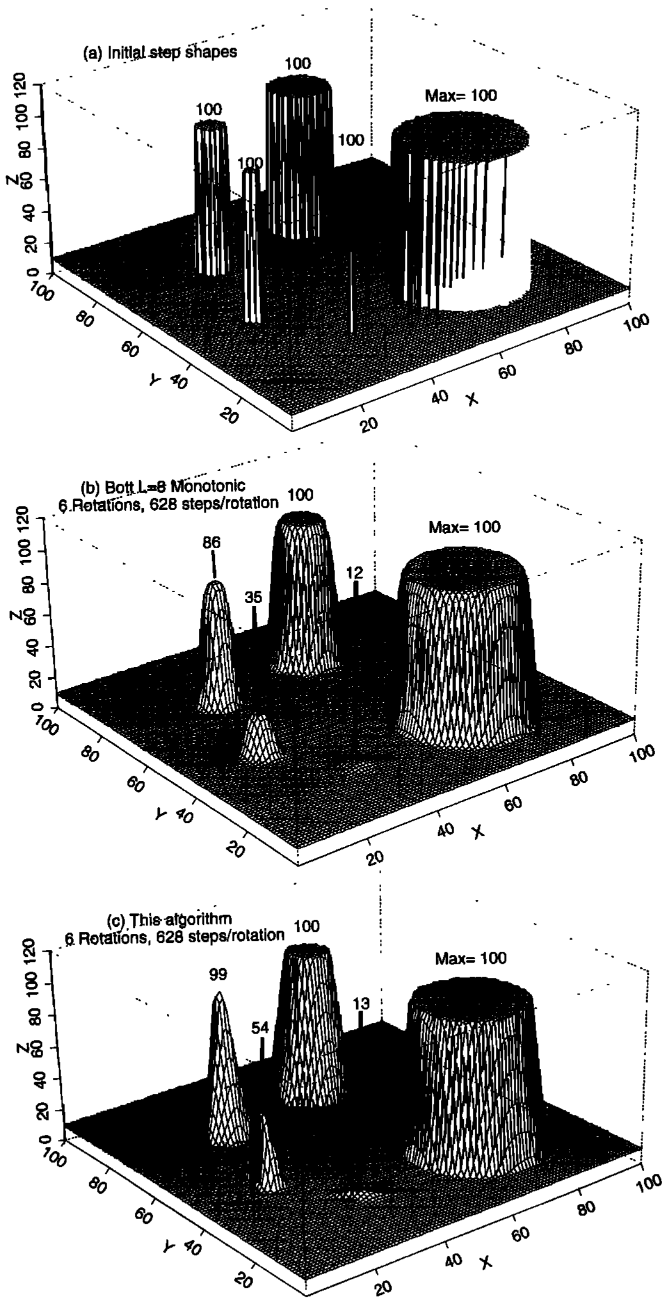


Figure 4. Same as Figure 3 except for a cylinder or "top hat" shape. (a) Initial and exact tracer distribution. (b) Distributions calculated using *Bott* [1992] L=8. (c) Distributions calculated using this algorithm.

severe test of rotating a "top hat" or cylinder distribution is shown in Figure 4. For poorly resolved cylinders this scheme is significantly better than the highest-order Bott approach. Even for the well-resolved cylinders, the cylinder edges appear more distinct and are less rounded than the L=8 algorithm.

Figure 5 shows the advection of a more smoothly varying cosine hill and shows a minor shortcoming of this algorithm: the slight aggregation of mass around the peaks of well-resolved tracer features. While both the *Bott* [1992] algorithm and this algorithm produce a slightly more flattened peak relative to the exact shape, the Bott approach flattens out ~6% below the true maximum, while this scheme flattens out within a fraction of a percent of the exact peak value. For well-resolved, smoothly varying shapes like this cosine hill, this relatively minor error slowly appears after long advection times. In contrast, for poorly resolved cosine hill shapes, this scheme is about as accurate as or even significantly more accurate than the Bott or *Prather* [1986] approach according to several error measures.

The rotating plane tests were performed using the finite-element *Toon et al.* [1988] technique, other orders of the *Bott* [1992] polynomial approach, the *Prather* [1986] method, and several versions of *Smolarkiewicz* [1983] successive correction method. Although calculated tracer distributions are not shown, error measures for these additional two dimensional tests are summarized in Figure 6. Error measures for each shape

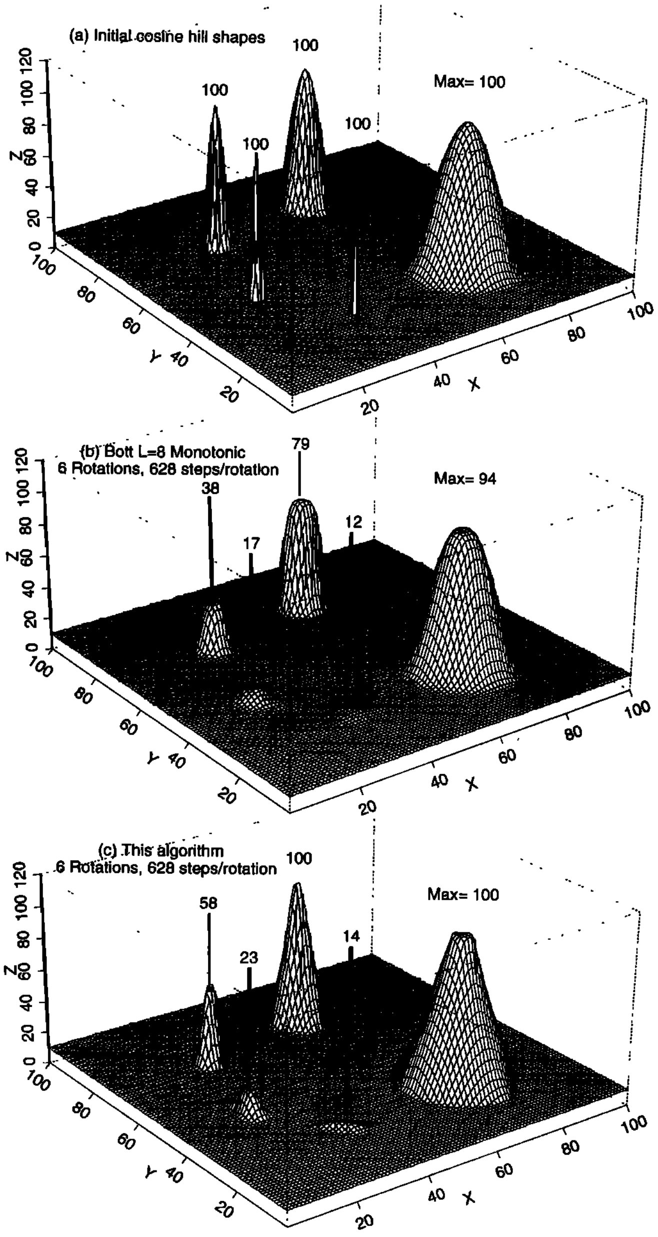


Figure 5. Same as Figure 3 except a cosine hill shape. (a) Initial and exact tracer distribution. (b) Distributions calculated using *Bott* [1992] L=8. (c) Distributions calculated using this algorithm.

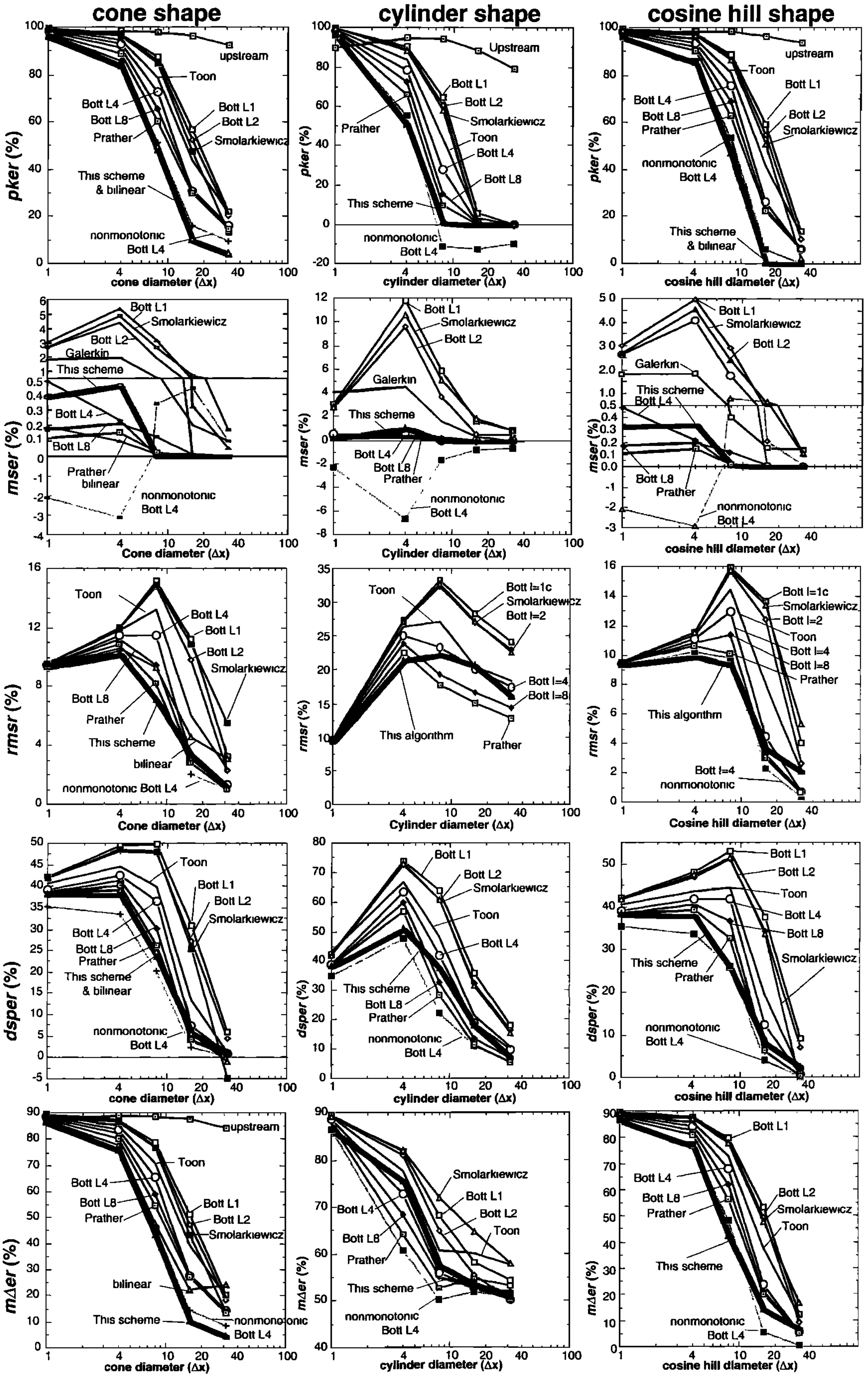


Figure 6. Error measures quantifying performance of this algorithm and algorithms by *Bott* [1992], *Smolarkiewicz and Grabowski* [1990], *Toon et al.* [1988], and *Prather* [1986] for 2-D rotating plane advection tests shown in Figures 3-5. (left) Cone shape. (middle) Cylinder shape. (right) Cosine hill shape. Rows from top to bottom refer to *pker* (equation (11)), *mser* (equation (12)), *rmser* (equation (13)), *dsper* (equation (14)), and *mΔer* (equation (15)).

are calculated over a circular disc subdomain with a radius that includes the initial shape plus $5\Delta x$ outside the edge of the initial shape. Thus errors for the $1\Delta x$ shape are averaged over all cell centers that are within a radius of $6(1+5)\Delta x$ (109 cells) closest to the center of the initial shape. For each larger shape, errors are averaged over the 149, 253, 529, and 1373 cells closest to the center of each shape. Minimum errors are usually zero, except for the nonmonotonic schemes, or slightly above zero, owing to numerical diffusion, and are not shown for brevity.

The top row of Figure 6 shows peak errors (*pker*) as a function of the size of the advected feature for many algorithms. Clearly, this scheme produces the best peak preservation of all algorithms tested, with the relative improvement most noticeable for features in the size range $4\text{-}16\Delta x$, where it is not unusual to see peak errors that are one-half to one-third the corresponding errors of the highest-order $L=8$ approach.

The second row of Figure 6 shows mass errors (*mser*). Note the expanded scale for mass errors near zero to highlight the differences between some schemes with low mass errors. While all schemes conserve mass over the entire 100x100 domain, not all the mass was within five grid cells of its correct location. For the monotonic schemes this "leakage" is mostly numerical diffusion. However, this mass error is also calculated for one nonmonotonic scheme, the *Bott* [1992] $L=4$ approach, and as shown in the second row of Figure 6, the higher-order nonmonotonic Bott scheme contains mass errors of up to -8% for the cylinder that is $4\Delta x$ in diameter, meaning that this nonmonotonic scheme artificially adds 8% of the initial mass to the domain within $5\Delta x$ of the edge of the proper tracer location.

The last three rows of Figure 6 show the root-mean-square (*rmser*), dispersion (*dsper*), and maximum absolute deviation (*mΔer*) errors. According to these performance measures, this scheme is usually slightly more accurate than the *Bott* [1992] $L=8$ or *Prather* [1986] approach, and for the larger cosine hill shapes, rms and dispersion errors are small but slightly worse than the best schemes tested here.

3.4. Shortcomings of Nonmonotonic Schemes

For some tests, nonmonotonic advection algorithms can yield very good scores of advection performance. Note in Table 1 and Figure 6 that the nonmonotonic *Bott* [1989b] scheme can often outscore all the algorithms tested here (e.g., note the dispersion errors for the cosine hill shape in Figure 6). However, such a good performance must be viewed with caution. To further highlight the shortcomings of nonmonotonic schemes, the rotating "top hat" cylinder test shown in Figure 4 is repeated for three nonmonotonic schemes and results are shown in Figure 7. Figure 7a shows that the *Toon et al.* [1988] technique contains appreciable noise throughout the background. The smallest two features are indistinguishable from the nonmonotonic oscillations generated by the scheme, and the larger advected shapes are poorly represented, with considerable nonphysical overshooting and undershooting. While poorly resolved shapes are calculated relatively well by the nonmonotonic *Bott* $L=8$ scheme (Figure 7b), the scheme contains appreciable overshooting and undershooting near the edges of the larger cylinder. By comparing the monotonic version of Bott $L=8$ (Figure 4b) with Figure 7b (nonmonotonic version), one notes the appreciable diffusion introduced when monotonic constraints are imposed, especially for poorly resolved features. Note, for example, the peak value of the second smallest cylinder falling from an extremely good value of 95 (Figure 7b) for the nonmonotonic solution to 35 (Figure 4b) when monotonic constraints are imposed. For this particular size cylinder the nonmonotonic cylinder with the very good calculated peak contains 3% more tracer mass within $5\Delta x$ radial distance beyond the edge of the initial shape. Thus tracer mass is not locally conserved.

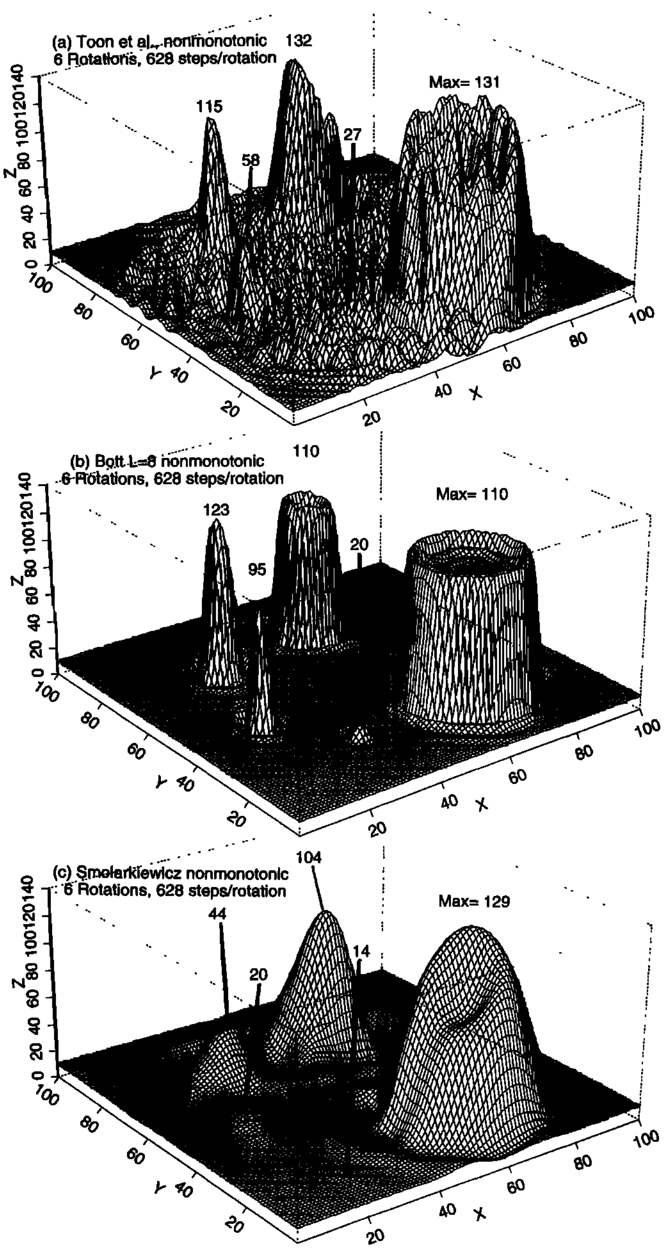


Figure 7. Same as Figure 4, except for calculations using nonmonotonic algorithms: (a) Galerkin finite element [*Toon et al.*, 1988], (b) *Bott* [1992] $L=8$, and (c) *Smolarkiewicz* [1983] (three corrective steps).

The *Smolarkiewicz* [1983] technique shown in Figure 7c produces appreciable unrealistic deformations of the initial tracer field for all sizes of this tracer distribution and significant overshooting for the large shapes, and the smaller shapes are virtually indistinguishable from the smoothly varying background nonmonotonic noise generated by the scheme. For all the nonmonotonic schemes shown in Figure 7, mass is

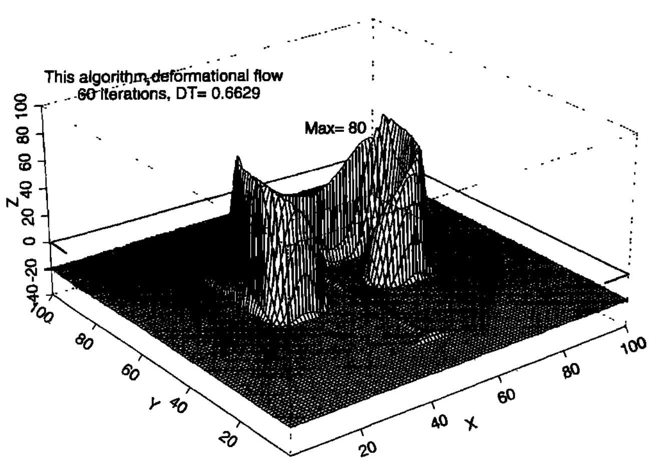


Figure 8. Advection of a 15Δx cone in deformational flows after 60 time steps using time step DT=0.6629.

not locally conserved around each shape. The *Toon et al.* [1988] and *Smolarkiewicz* [1983] schemes contain mass errors of up to 13% for the shapes shown in Figure 7.

Under some conditions, nonmonotonic algorithms can produce relatively good advection scores. Unfortunately, Figure 7 clearly shows that nonmonotonic schemes can produce appreciable mass errors or other highly undesirable distortions when poorly resolved features or steep gradients are present in the advected tracer fields. In recent literature, many nonmonotonic schemes yield extremely low error scores [e.g., *Jakob-Chien et al.,* 1995; *Li and Chang,* 1996; *Chlond,* 1994; *Toon et al.,* 1988], but these schemes produce unrealistic results if subjected to conditions similar those shown in Figure 7, and nonmonotonic algorithms contain nonnegligible mass imbalances locally around advected shapes.

3.5. Deformational Flow Test

Here the locally deformational flow test described by *Bott* [1989a] is repeated. The x velocity u=sin[πi/25]sin[πj/25], and the y velocity v=cos[π(i - 0.5)/25]cos[π(j+.5)/25], where i and j refer to the x and y grid indices numbered from the lower left corner of a 100x100 grid. A 15Δx radius cone is initially centered in the domain, and here a negative background (-20) and a cone peak mixing ratio of 100 are used to show how this scheme can accurately advect features with nonzero background and negative values without modification. Figure 8 shows the calculated shape after 60 time steps. For this simulation, Δt = 0.6629, and Figure 8 can be compared with the exact shape shown in Figure 3c of *Staniforth et al.* [1987]. Even at this time, the exact shape is sheared into curtains of tracer that are thinner than the resolution of this test, but the oscillation-free background and the very good shape and peak preservation of this scheme are evident in Figure 8.

3.6. Atmospheric transport

Figure 9 shows results of this scheme applied to realistic time- and density-varying atmospheric flows, where an ozone-like tracer that is chemically produced in the lower troposphere over two urban areas is transported by ambient winds. Three-dimensional velocity fields are generated and stored at hourly intervals throughout a 5-day period using the National Center for Atmospheric Research (NCAR) mesoscale meteorology model. Both the meteorology model and the tracer advection model use a horizontal resolution of 80 km and 15 vertical sigma pressure layers between the surface and the 100 mbars pressure level, and the domain shown in Figure 9 is horizontally subdivided into a grid of 35x38 cells. The time step for this test is 1200 s, which is close to a Courant number limit for the high winds encountered near the jet stream in the upper portions of the domain. An ozone-like constituent with a background mixing ratio of 50 ppb everywhere is continuously produced at a rate of 5 ppb h^{-1} in the lowest five model layers from 1000 to 900 mbars, which approximately represents a planetary boundary layer depth. The chemical formation occurs only in the small 2x2 Δx^2 (Atlanta) or 1x2 Δx^2 (Chicago) shaded grid cells shown in Figure 9. The tracer distributions and wind fields in the layer 1000-900 mbars are shown after 4.5 days of continuous chemical formation. Wind fields show a midlatitude cyclonic storm traversing the domain from west to east, and at the lowest atmospheric layers the ozone-like tracer produced over each urban area is advected into the center of the eastward moving low pressure convergent area and lifted up by the converging winds for transport above the boundary layer eastward by higher winds aloft. For this simulation the background constant mixing ratio is perfectly maintained, and the multiple peaks in concentration result from dynamical transport processes.

4. Discussion

Figures 2-8 demonstrate that significant improvement of a simple, low-order advection algorithm can be achieved by very slightly aggregating mass near local extremes of a tracer distribution during flux calculations. For a wide range of test

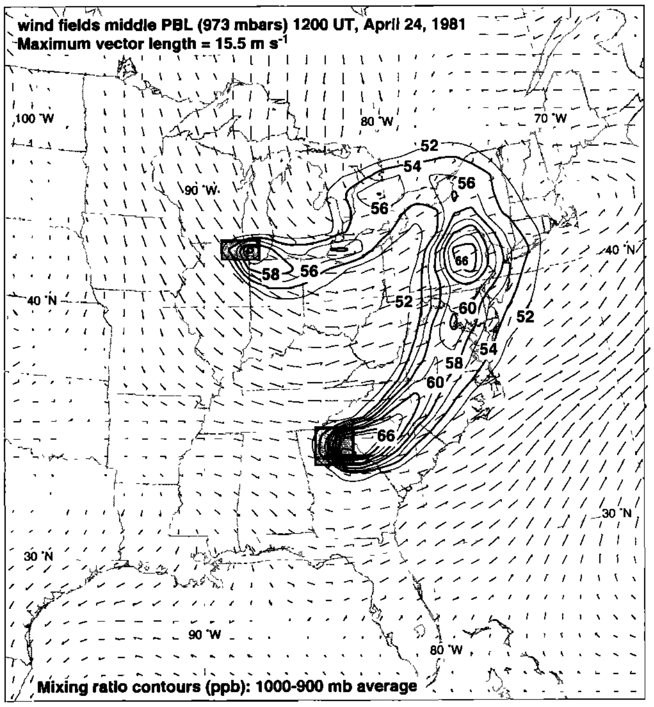

wind fields middle PBL (973 mbars) 1200 UT, April 24, 1981
Maximum vector length = 15.5 m s-1
100 °W
80 °W
70 °W
90 °W
52
54
56
56
58
56
66
52
60
52
54
58
60
66
40 °N
40 °N
30 °N
30 °N
90 °W
80 °W
Mixing ratio contours (ppb): 1000-900 mb average

Figure 9. Wind fields and mixing ratio (ppb) of an ozone-like tracer chemically formed continuously at 5 ppb h^{-1} between 1000 and 900 mbars over two urban areas (shown as shaded boxes) and advected by time-varying winds for 4.5 days. Tracer distributions are averaged over five model layers between 1000 and 900 mbars. Outside the chemical formation area, a constant 50 ppb was initially specified.

problems, particularly when tracer shapes are resolved by fewer than 10-20 grid cells, this algorithm produces errors that are appreciably smaller than errors calculated using higher-order algorithms that are appreciably more complex and computationally expensive.

4.1. Flux Adjustment Near Extremes

The use of empirically derived factors to enhance or "optimize" the performance of advection schemes is not unusual. For example, *Smolarkiewicz* [1983] suggests "tuning" an antidiffusion velocity factor *Sc* to reduce errors and recommends increasing antidiffusive velocities by 1.06 for one 2-D test, but suggests a range 1-1.08 for other applications. Similarly, *Takacs* [1985] empirically adjusts another α factor (different from the α here) to minimize errors when advecting a triangular distribution in one dimension. *Yamartino* [1993] optimizes two adjustable factors to minimize errors for a single 2-D test. A "steepening factor" which is invoked under empirically derived discontinuity conditions is also incorporated into the piecewise parabolic method (PPM) described by *Carpenter et al.* [1990]. Likewise, *Chlond* [1994] applies an elaborate exponential flux algorithm near "danger zones" that are identified using empirically defined thresholds for several indicator parameters that identify tracer regions that are susceptible to undesirable behavior.

The Courant-number-dependent formulations used here to "aggregate mass" near extremes to reduce numerical diffusion (equation 10) are optimized as described in the appendix for a wide range of test problems. As described by *Walcek and Aleksic* [1998], the quasi-empirical adjustment of a low-order advective flux calculation near extremes essentially forces the low-order algorithm to mimic the desirable behavior of higher-order schemes near local extremes of a tracer distribution. Equations 10a-b were empirically derived by subjecting a wide variety of shapes in both one dimension and two dimensions to a wide number of combinations of sharpening factors upwind and downwind of local extremes over the Courant range 0-1, and choosing optimal values for α_u and α_d.

In this scheme the empirical adjustment is applied at only two cell edges upwind and downwind of each local extreme of a tracer distribution. For most modeling applications the number of cells defined as a "local extreme" is typically only a small fraction of the total number of cells in the domain, and monotonic limitations (equation (6)) often "check" the initial estimates of outflowing fluxes irrespective of location near peaks. Diagnostic analysis of this algorithm for the rotating plane experiments (Figures 3-5) showed that fluxes are only adjusted using (10) at 3-3.5% of cell edges encountered during the advection calculations. The average flux adjustment at these few cell interfaces resulting from using $\alpha > 1$ amounted to 2-5%. Empirical flux adjustments greater than 10% were only imposed at less than 0.9% (step shape) or 0.07% (cone or cosine shape) of the cell edges. However, as shown in Figure 6, these relatively minor adjustments at an extremely small fraction of the grid cells removed 20-80% of the errors inherent by using $\alpha=1$ everywhere and ignoring (10).

4.2. Computational requirements

This algorithm is computationally simpler than other higher-order schemes. The method must be invoked in a flow-direction-dependent sequence, and local extremes must be identified. These aspects require several "if" tests for each grid cell, but the number of multiplication and additions required per updated mixing ratio is appreciably smaller than most higher-order advection algorithms. Table 2 shows the relative computational time of the 2-D tests considered here. Also shown is the mean of the rms errors (from Figure 6) for the rotating cone test (Figure 3). For slightly less than 3 times the computational cost of upstream advection, one obtains advection performance that is significantly superior to the highest-order schemes considered here. The only algorithms tested here that were computationally cheaper were algorithms containing 38-164% greater rms errors for this test, and the only algorithms that contained rms errors that were slightly worse than this algorithm required 2-3 times the computational expense.

Table 2. Computational Time

Algorithm	Relative Computation Time	Mean rms Errors for Figure 3 Cones, %
Upstream	1	18.4
Bott [1992] L=1	2.2	10.2
Bott [1992] L=2	2.5	9.7
Toon et al. [1988]	2.8	8.6
This algorithm	2.9	6.2
Bott [1992] L=4	3.1	7.4
Smolarkiewicz [1983]	4.0	10.5
Bott [1992] L=8	5.1	6.8
Prather [1986]	6.6	6.4

Figure 10 shows a one-page FORTRAN program containing a generally applicable subroutine that can be utilized for other applications. The program calculates two-dimensional advection by sequentially calling the 1-D subroutine in each dimension. In this program, rotating plane velocities (section 3.3) or the deformational flows (section 3.5) are defined in the main program. The 1-D advection subroutine has been implemented unchanged into several models of various coordinate systems and domains for calculating tracer advection. For at least several years after publication the author will provide a highly commented version of this program together with sample input and output to users via electronic mail requests to walcek@asrc.cestm.albany.edu.

Appendix: Derivation of Equation (10)

The adjustment of factor α to improve advection performance was initially investigated by *Sweby* [1984] and others, who found that there are advantages to setting $\alpha>1$ or setting α to be functionally related to the surrounding tracer distribution [e.g., *Harten and Osher,* 1987; *Thuburn,* 1996]. Monotonic solutions can be obtained through the appropriate choice of α, and α adjustments that induce "compression" [*Sweby,* 1984], "slope mismatch" [*Lin et al.*, 1994], or "steepening" [*Walcek and Aleksic,* 1998] were found to vastly improve advection performance for many test problems. However, these researchers found that tracer mass tends to aggregate around local peaks when using $\alpha>1$ over large areas of a tracer distribution, which may be desirable for advecting small features but produces unrealistic solutions when applied to tracer distributions containing "smooth" peaks or well-resolved extremes. Thus a triangle-like distribution is quickly transformed into a "plateau" peak if $\alpha>1$ is used everywhere [e.g., see *Walcek and Aleksic,* 1998, Figure 4]. Walcek and Aleksic found that this "peak aggregation" problem can be largely overcome by severely limiting the number of cells

```
        PROGRAM  ADVECT2D
C This code advects a 15DX cone-shaped mixing ratio distribution in two
C   dimensiona by sequentially advecting x and y directions calling 1-D
C   subroutine each dimension. Equation #s refer to manuscript.
C
      PARAMETER(IMAX=100,JMAX=100,IX=110,IX1=IX+1)
      DIMENSION VC2D(IMAX,JMAX),U2D(0:IMAX,JMAX),V2D(IMAX,0:JMAX),
     &   V0(IMAX,JMAX)
      COMMON/sub/Q0(0:IX1),QN(IX),U(0:IX),DEN0(IX),DEN1(IX),DT,DXX(IX),
     &           DD0(0:IX)
      DATA AI0S,AJ0S, DX, UMX,VCBACK,VCSTEP,PII, D0/
     &     25.,  50.,  1., 1.,-20.,  120., ,3.141592653589793,1. /
      DATA AN,SUM2,UMAX,RMSER,VCMAX,VCMIN/4*0.,-1.E10,1.E10/
C
      WRITE(6,*) ' Wind fields?  0-rotating;    or  1-divergent winds'
      READ(5,*) IWNDTY
      IF(IWNDTY.EQ.1) AI0S= 50.5
      ILOP= AI0S-21.9                   ! needed for printouts
C
C Define wind fields (rotation or divergent) and initial mixing ratios
C  Cone at (25,50) for rotating winds; Cone at (50,50) divergent winds
      DO 10 J=JMAX,0,-1
      DO 11 I=0,IMAX
      IF(J.GT.0) U2D(I,J)=-2.*UMX*(REAL(J)-REAL(JMAX)/2.-.5)/REAL(JMAX)
      IF(I.GT.0) V2D(I,J)= 2.*UMX*(REAL(I)-REAL(IMAX)/2.-.5)/REAL(IMAX)
      IF(IWNDTY.EQ.1) THEN
       XA=PII/25.
       IF(J.GT.0) U2D(I,J)=UMX*SIN(XA*REAL(I))*SIN(XA*(REAL(J)))
       IF(I.GT.0) V2D(I,J)=UMX*COS(XA*(REAL(I)-.5))*COS(XA*(REAL(J)+.5))
      END IF
      IF(I.EQ.0.OR.J.EQ.0) GOTO 11
      DXX(J)= DX
      UMAX= MAX(ABS(U2D(I,J)),ABS(V2D(I,J)),UMAX)
      RX= SQRT((REAL(I)-AI0S)**2.+(REAL(J)-AJ0S)**2.)
      VC2D(I,J)= VCBACK + VCSTEP*MAX( 0.,  1.-RX/15.)
      V0(I,J)= VC2D(I,J)
   11 CONTINUE
   10 IF(J.GT.0) WRITE(6,100) J,(NINT(VC2D(I,J)),I=ILOP,ILOP+42)
C
  100 FORMAT(' ',44I3)
      IF(IWNDTY.EQ.0) THEN
       WRITE(6,*) ' #iterations per revolution? must be>',PII*(IMAX-1)
       WRITE(6,*) '    Followed by the number of revolutions desired'
       READ(5,*) NITPREV, ANREV
       DT= PII*REAL(IMAX-1)*DX/UMAX/REAL(NITPREV)
       NSTEP= REAL(NITPREV)*ANREV + .9999
      ELSE
       WRITE(6,*) ' Courant #?  eg. C=.6629 (to match Staniforth figs)'
       WRITE(6,*) '  Followed by # of iterations? e.g. 20,40,60,80,400'
       READ(5,*) CURNT, NSTEP
       DT= CURNT*DX/UMAX
      END IF
      WRITE(6,*) ' Max wind speed, Courant number=',UMAX,UMAX*DT/DX
      Q0(0)= VCBACK       ! boundary conditions: lateral inflow/outflow
      Q0(IMAX+1)= VCBACK
C
      DO 30 IT=1,NSTEP             !  Beginning of time iteration loop
      IF(MOD(IT,10).EQ.0)WRITE(6,*)' processing iteration',IT,' -',IT+10
C
      DO 20 J=1,JMAX          ! Do I-advection first over all J-lines
        DO 22 I=1,IMAX
         DEN0(I)= D0                                               !Eq-5
         DEN1(I)= D0-DT/DXX(I)*(D0*U2D(I,J)- D0*U2D(I-1,J  ))      !Eq-5
         DD0(I)=DEN0(I)
         Q0(I)= VC2D(I,J)
   22    U(I)= U2D(I,J)
        DD0(0)= D0
        U(0)= U2D(0,J)
         CALL  ADVEC1D(IMAX)
        DO 20 I=1,IMAX
   20 VC2D(I,J)= QN(I)
C
      DO 30 I=1,IMAX          !  Do J-advection next over all I lines
        DO 32 J=1,JMAX
         Q0(J)= VC2D(I,J)
         DEN0(J)= D0-     DT/DXX(I)*(D0*U2D(I,J)-D0*U2D(I-1,J  )) !Eq-5
         DEN1(J)= DEN0(J)-DT/DXX(J)*(D0*V2D(I,J)-D0*V2D(  I,J-1)) !Eq-5
         DD0(J)= D0
   32    U(J)= V2D(I,J)
        DD0(0)= D0
        U(0)= V2D(I,0)
         CALL  ADVEC1D(JMAX)
        DO 30 J=1,JMAX
   30 VC2D(I,J)= QN(J)            ! End of time and I iteration loop
C
C  Print out final distributions, assess errors near initial shape
      WRITE(6,*) ' mixing ratio at N=',NSTEP
      DO 50 J=JMAX,1,-1
      DO 55 I=1,IMAX
       RX= SQRT((REAL(I)-AI0S)**2.+(REAL(J)-AJ0S)**2.)
       IF(RX.LT.20) RMSER= RMSER + (VC2D(I,J)-V0(I,J))**2.    !Eq-13
       IF(RX.LT.20) AN= AN + 1.
   55 VCMAX= MAX( VC2D(I,J),VCMAX )
   50 WRITE(6,100) J,(NINT(VC2D(I,J)),I=ILOP,ILOP+42)
      RMSER=  100.*SQRT( RMSER/AN)/VCSTEP                    !Eq-13
      PEAKER= 100.*(VCBACK+VCSTEP-VCMAX)/VCSTEP              !Eq-11
      WRITE(6,245) NSTEP,VCMAX
      IF(IWNDTY.EQ.0) WRITE(6,246) RMSER,PEAKER
      STOP
  245 FORMAT(' after',I6,' iterations',/,F10.4,' Maximum mixing ratio')
  246 FORMAT(' ', f12.5,'% RMS error', /,F12.5,'% peak error')
      END
```

Generally applicable 1-D subroutine

```
      SUBROUTINE ADVEC1D(IDIM)
C This subroutine calculates change in mixing ratio (Q0) during time
C  step DT due to advection along a grid IDIM in length. Mixing ratios
C  from host code (C) are loaded into Q0 array, which is updated to QN.
C  Velocities (U) and fluxes (FLUX) are specified at cell FACES, having
C  dimensions 0:IDIM. U, Q0, QN, DXX and FLUX indices defined here:
C  Fluid densities flowing across each face (DD0), & beginning and end
C  of each dimension step (DEN0, DEN1) are defined in HOST CODE
C
C I grid->    |  1  |  2  |  I-1  |   I  |..   ..|  IDIM  | <- host grid
C U-array-> u(0)  u(1)  u(2)   u(i-1)  u(i)           u(IDIM)
C Flux -> FLUX(0)              FLUX(I-1) FLUX(I)     FLUX(IDIM)
C DD0  ->  DD0(0)               DD0(I-1)  DD0(I)      DD0(IDIM)
C C-array->   | C(1)| C(2)| C(I-1)| C(I) |..   ..| C(IDIM)| mixing ratio
C DXX-arry->  | Dx1 | Dx2 | DxI-1 | DxI  |..   ..| DxIDIM |
C Density->   | Dd1 | Dd2 | DdI-1 | DdI  |..   ..| DdIDIM |
C                  Q0 defined along 0 - IDIM+1 cells:
C      |        | QN  | QN  |  QN   |  QN  |       |  QN    |        |
C      |    Q0--|-Q0--|-Q0--|--Q0 --|--Q0--|..   ..|-- Q0 --|--Q0    |
C      |    0   | 1   |  2  | I-1   |  I   |       | IDIM   | IDIM+1 |
C    lower BC |            <---   Q0 grid   --->            | upper BC
C             Boundary conditions are stored in Q0 cells 0 & IDIM+1
      PARAMETER(IX=110,IX1=IX+1)
C  Input/output information passed to subroutine in common 'sub'
      COMMON/sub/Q0(0:IX1),QN(IX),U(0:IX),DEN0(IX),DEN1(IX),DT,DXX(IX),
     &           DD0(0:IX)
      DIMENSION FLUX(0:IX),VCMAX(IX),VCMIN(IX)
      LOGICAL IMXMN(0:IX1)
      DATA ZR0,IMXMN/0.,112*.FALSE./
C
      IMXMN(IDIM+1)=.FALSE.
C Identify local max and min, specify mixing ratio limits at new time
      DO 5 I=1,IDIM
        IMXMN(I)=Q0(I).GE.MAX(Q0(I-1),Q0(I+1)) .OR.  !=true if local
     &           Q0(I).LE.MIN(Q0(I-1),Q0(I+1))       !   extreme at i
        CK1= Q0(I)
        CK2= Q0(I)
        IF(U(I  ).LT.ZR0) CK1= Q0(I+1)
        IF(U(I-1).GE.ZR0) CK2= Q0(I-1)
        VCMAX(I)= MAX( Q0(I), CK1, CK2 )                 ! Eq. 7
    5   VCMIN(I)= MIN( Q0(I), CK1, CK2 )                 ! Eq. 7
C
C Update mixing ratios and limit Fluxes going UP where u>0
      IF(U(0).GE.ZR0) FLUX(0)=Q0(0)*U(0)*DT*DD0(0) !upstream at boundary
      DO 10 I=1,IDIM
      IF(U(I).LT.ZR0) GOTO 10
      IF(U(I-1).LT.ZR0) THEN              ! outflow-only cell: upstream
        FLUX(I)= Q0(I)*U(I)*DT*DD0(I)
       ELSE
        X1= DT*U(I)/DXX(I)                ! Courant number
        X1N= (1.-X1)*(Q0(I+1)-Q0(I-1))/4.
        CF= Q0(I) + X1N                                        !Eq-4a
        IF(IMXMN(I-1)) CF= Q0(I) +MAX(1.5,1.2  +.6 *X1)*X1N   !Eq-10b
        IF(IMXMN(I+1)) CF= Q0(I) +        (1.75 -.45*X1)*X1N   !Eq-10a
        CF1= MIN( MAX( CF, MIN(Q0(I),Q0(I+1))  ), MAX(Q0(I),Q0(I+1)) )
        QN(I)= MAX(  VCMIN(I),   MIN(   VCMAX(I),              !Eq-3 & 8
     &  (Q0(I)*DEN0(I)-X1*CF1*DD0(I)+FLUX(I-1)/DXX(I))/DEN1(I)  ))
        FLUX(I)= DXX(I)*(Q0(I)*DEN0(I) - QN(I)*DEN1(I)) + FLUX(I-1)
      END IF                                                   !Eq-9a
   10 CONTINUE
C
C Update mixing ratios and limit Fluxes going DOWN where u<0
      IF(U(IDIM).LT.ZR0) FLUX(IDIM)=Q0(IDIM+1)*U(IDIM)*DT*DD0(IDIM)
      DO 20 I=IDIM,1,-1                                        !Eq-13
      IF(U(I-1).GE.ZR0) THEN            ! Inflow-only cell
        IF(U(I).LT.ZR0) QN(I)=  MAX(  VCMIN(I),   MIN(   VCMAX(I),
     &   (Q0(I)*DEN0(I)-FLUX(I)/DXX(I) + FLUX(I-1)/DXX(I))/DEN1(I) ))
       ELSE
        X1=  DT*ABS(U(I-1))/DXX(I)      ! Courant number
        X1N= (1.-X1)*(Q0(I-1)-Q0(I+1))/4.
        CF= Q0(I) + X1N                                        !Eq-4b
        IF(IMXMN(I+1)) CF= Q0(I) +MAX(1.5,1.2  +.6 *X1)*X1N   !Eq-10b
        IF(IMXMN(I-1)) CF= Q0(I) +        (1.75 -.45*X1)*X1N   !Eq-10a
        CF1= MIN( MAX( CF, MIN(Q0(I),Q0(I-1)) ), MAX(Q0(I),Q0(I-1)) )
        IF(U(I).GE.ZR0) CF1= Q0(I)      ! outflow-only cell upstream
        QN(I)= MAX(  VCMIN(I),  MIN(   VCMAX(I),               !Eq-3 & 8
     &   (Q0(I)*DEN0(I)-FLUX(I)/DXX(I)-X1*CF1*DD0(I-1))/DEN1(I) ))
        FLUX(I-1)=DXX(I)*(QN(I)*DEN1(I) - Q0(I)*DEN0(I)) + FLUX(I)
      END IF                                                   !Eq-9b
   20 CONTINUE
      RETURN
      END
```

Figure 10. FORTRAN program with 1-D advection subroutine of this algorithm for 2-D tests.

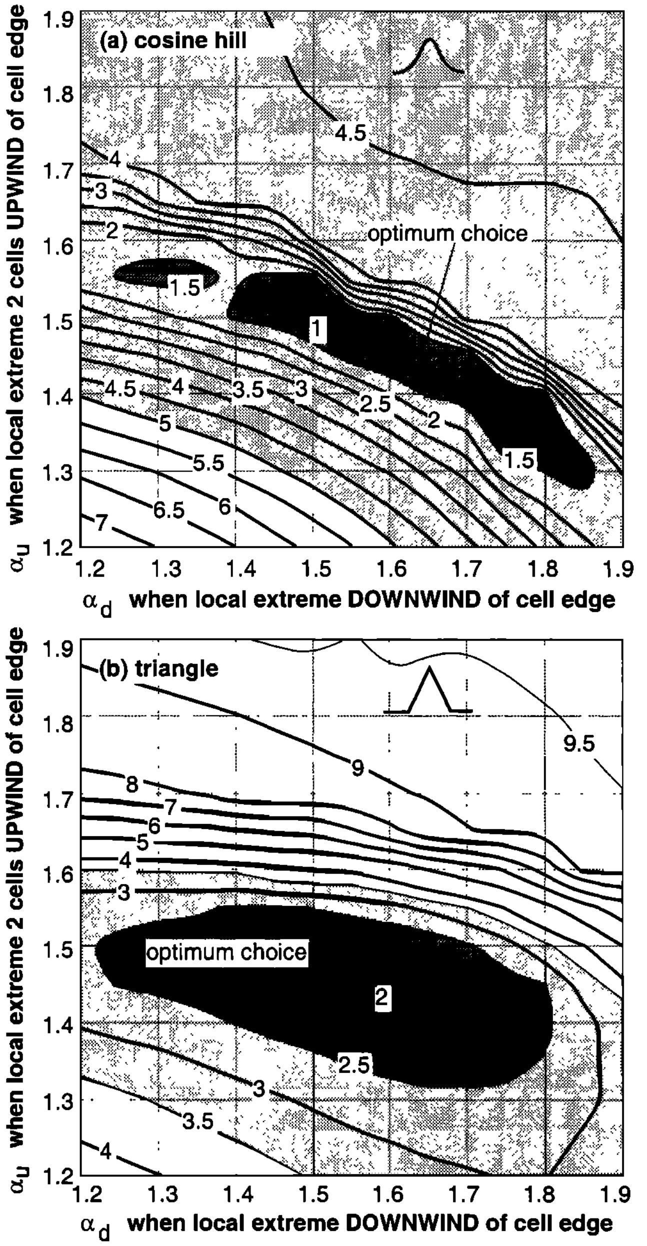


Figure 11. Root-mean-square errors (in percent) for advecting (a) cosine-shaped hill ($8\Delta x$ period) and (b) a triangle shape ($5\Delta x$ half width). Errors are given as a function of steepening factor α (from equation 10) applied upwind and downwind of local extremes. Shapes advected 50 cells at a Courant number 0.2994 (167 time steps). The rms errors are averaged over 40 cells surrounding the exact shape location. Shaded areas denote regimes where this scheme is more accurate than monotonic versions of higher-order schemes by *Bott* [1992] L=4 (light shading) or *Prather* [1986] (dark shading). Crosses denote "optimum" choices of α_u and α_d which produce the lowest rms error.

where "steepening" (using α>1) is applied. Here only two cells surrounding each local extreme of a tracer distribution are "steepened". Thus, for tracer features that are resolved by a large number of grid cells, the additional steepening influences only a small fraction of the tracer distribution. For poorly resolved tracer distributions, a proportionally larger fraction of the tracer shape experiences "aggregation" during advection updates. Thus the scheme represents a "hybrid" approach where the positive aspects of setting α>1 for poorly resolved shapes are maintained while the undesirable influences of setting α>1 for well-resolved peaks are limited.

The Courant-number-dependent equations used to "aggregate" mass around local extremes to significantly improve advection performance (equation (10a) and (10b)) were derived by performing numerous 1-D advection tests of a variety of tracer shapes over the Courant number range 0-1. The 14 shapes shown in Figure 5 of *Walcek and Aleksic* [1998] were advected 50 grid cells, requiring 56-500 time steps, depending on the Courant number tested. It was hypothesized that aggregation would be required upwind of and downwind of extremes, so for each advected shape rms errors were calculated for all possible combinations of α_u and α_d in the range $1 < \alpha < 2$. Figure 11 shows a contour diagram of the percent rms errors generated by this algorithm as a function of α_d (horizontal axis) and α_u (vertical axis) for a cosine hill (Figure 11a) and a triangle (Figure 11b) shape advected for 167 time steps at a Courant number 0.2994. The cosine hill is defined by

$$Q = 2 + \cos(\pi\Delta x/4) \quad |\Delta x|<4 ,$$

Q=1 everywhere else. The triangle is specified as

$$Q = 3 - 2|\Delta x|/5 \quad |\Delta x|<5,$$

Q=1 everywhere else. Here Δx is the number of cells between a given cell and the peak or center of the initial shape in the 1-D domain of 100 grid cells.

It is readily apparent in Figure 11 that there are optimal choices that minimize the rms advection errors, and the best choices for α_u and α_d are marked with a cross. Shaded areas in Figure 11 denote regimes where rms errors calculated using this algorithm are lower than the *Bott* [1992] L=4 (light shading) or the more accurate multimoment monotonic *Prather* [1986] (dark shading) schemes. While the optimum choice of α_u and α_d for one shape is not the best choice for the other

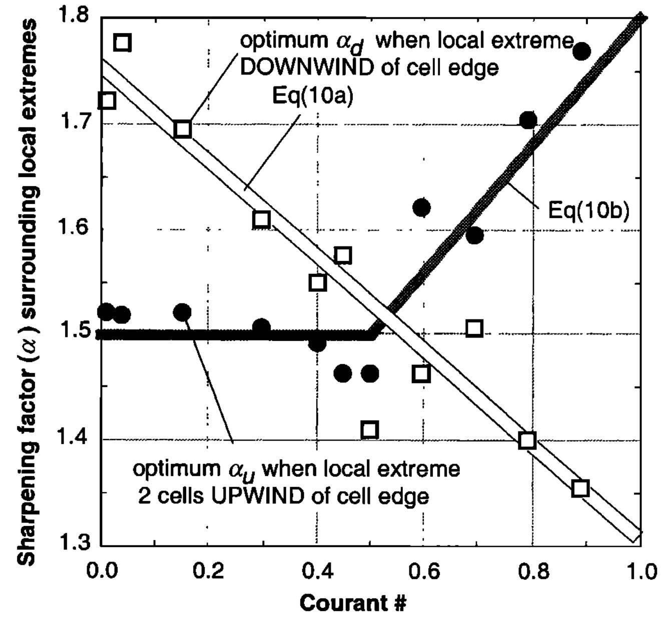


Figure 12. Optimum choice of aggregation factors for cell edges if a local extreme is downwind of a cell edge (open squares) and upwind of a cell edge (solid circles) as a function of Courant number. Each point is the average of the optimum factors for advecting 14 shapes of various sizes 50 cells in one dimension.

shape, one can choose values for α_u and α_d that produce better scores than both Prather and Bott for both shapes. For all the 1-D test shapes presented by *Walcek and Aleksic* [1998] the optimum α_u and α_d combination yielding the smallest rms error was identified for each shape at a particular Courant number. The additional test shapes include cosine or sine shapes of various periods (periods of $10\Delta x$ and $20\Delta x$), Gaussian [Q=1+2exp(-Δx^2/49)], exponential [Q=1+2exp(-|Δx|/7], circle [Q=1+ 2(1-Δx^2/64)$^{0.5}$, |Δx|<8], parabola [Q=3-2Δx^2/49, |Δx|<7], triangles (half-width $10\Delta x$), steps, and "tooths" [Q=1+2(1-Δx/7) for 0<Δx<7 only].

At each Courant number the optimum α_u and α_d were averaged for all shapes tested, and these averages are plotted on Figure 12. It is readily apparent that there is a Courant-number-dependent trend through the points plotted. Equations (10a) and (10b) represent a rough fit through the optimum values shown in Figure 12, although there are an appreciable range and scatter of the optimum α_u and α_d values that can be used that will make this algorithm better than existing higher-order schemes. It was found that generally, using α slightly less than (10) worked better for "smoothly varying" tracer distributions, while using α slightly greater than (10) improved error scores for poorly resolved tracer distributions relative to the results shown here. Using (10) around extremes represents a reasonable "average" that produces fairly good results that are nearly always better than the *Bott* [1992] L=4 scheme and often better than results of *Prather* [1986], especially if advected features are resolved by fewer than 8-12 grid cells. In the limit of advecting extremely small features, *Walcek and Aleksic* [1998] found that a $2\Delta x$-wide step function could be advected perfectly for any Courant number for constant 1-D winds by using α>4 everywhere, which also holds for this algorithm, although using α>4 everywhere leads to unrealistic aggregation of mass around local extremes for large tracers of any shape.

Acknowledgments. Eric Berge of the Norwegian Meteorological Institute provided valuable motivation and insights during the preparation of this manuscript. The following organizations supported this research: the U. S. EPA (grant R827929010), the U. S. AFOSR (grant F4962 092J0018), and NASA (grant NAG5-2714).

References

Bott, A., A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes, *Mon. Weather Rev. 117*, 1006-1015, 1989a.

Bott, A., Reply to comment by P. K. Smolarkiewicz on "A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes" by A. Bott, *Mon. Weather Rev., 117*, 2633-2636, 1989b.

Bott, A., Monotone flux limitation in the area-preserving flux-form advection algorithm, *Mon. Weather Rev. 120*, 2592-2602, 1992.

Bott, A., The monotone area-preserving flux-form advection algorithm: Reducing the time-splitting error in two-dimensional flow fields, *Mon. Weather Rev., 121*, 2637-2641,1993.

Carpenter, R. L., K. K. Droegemeier, P. R. Woodward, and C. E. Hane, Application of the piecewise parabolic method (PPM) to meteorological modeling, *Mon. Weather Rev., 118*, 586-612, 1990.

Chlond, A., Locally modified version of Bott's advection scheme, *Mon. Weather Rev., 122*, 111-125, 1994.

Chock, D. P., A comparison of numerical methods for solving the advection equation, III, *Atmos. Environ., 25A*, 853-871, 1991.

Clappier, A., A correction method for use in multidimensional time-splitting advection algorithms: Application to two- and three-dimensional transport, *Mon. Weather Rev., 126*, 232-242, 1998.

Colella, P., and P. R. Woodward, The piecewise parabolic method (PPM) for gas-dynamical simulations, *J. Comp. Phys., 54*, 174-201, 1984.

Costa, A. A., and A. J.-C. Sampaio, Bott's area-preserving flux-form advection algorithm: Extension to higher orders and additional tests, *Mon. Weather Rev., 125*, 1983-1988, 1997.

Dabdub, D., and J. H. Seinfeld, Numerical advective schemes used in air quality models: Sequential and parallel implementation, *Atmos. Environ., 28*, 3369-3385, 1994.

Easter, R. C., Two modified versions of Bott's positive-definite numerical advection scheme, *Mon. Weather Rev., 121*, 297-304, 1993.

Gasdag, J., Numerical convective schemes based on accurate computation of space derivatives, *J. Comput. Phys., 13*, 100-113, 1973.

Harten, A., and S. Osher, Uniformly high-order accurate nonoscillatory schemes, I, *SIAM J. Numer. Anal., 24*, 279-309, 1987.

Jakob-Chien, R., J. J. Hack, and D. L. Williamson, Spectral transform solutions to the shallow water test, *J. Comput. Phys., 119*, 164-187, 1995.

Li, Y., and J. S. Chang, A mass conservative positive definite and efficient Eulerian advection scheme in spherical geometry and on a nonuniform grid, *J. Appl. Met., 35*, 1897-1913, 1996.

Lin, S.-J., and R. B. Rood, Multidimensional flux-form semi-Lagrangian transport schemes, *Mon. Weather Rev., 124*, 2046-2070, 1996.

Lin, S.-J, W. C. Chao, Y. C. Sud, and G. K. Walker, A class of the van Leer-type transport schemes and its application to the moisture transport in a general circulation model, *Mon. Weather Rev., 122*, 1573-1593, 1994.

Pietrzak, J., The use of TVD limiters for forward-in-time upstream-biased advection schemes in ocean modeling, *Mon. Weather Rev., 126*, 812-830, 1998.

Prather, M. J., Numerical advection by conservation of second-order moments, *J. Geophys. Res., 91*, 6671-6681, 1986.

Rasch, P. J., Conservative shape-preserving two-dimensional transport on a spherical grid, *Mon. Weather Rev., 122*, 1337-1350, 1994.

Rood, R. B., Numerical advection algorithms and their role in atmospheric transport and chemistry models, *Rev. Geophys., 25*, 71-100, 1987.

Russell, G.-L., and J. A. Lerner, A new finite-differencing scheme for the tracer transport equation, *J. Appl. Met., 20*, 1483-1499, 1981.

Schneider, H. R., A numerical transport scheme that avoids negative mixing ratios, *Mon. Weather Rev., 112*, 1206-1217, 1984.

Smolarkiewicz, P. K., A simple positive definite advection scheme with small implicit diffusion, *Mon. Weather Rev., 111*, 479-486, 1983.

Smolarkiewicz, P. K., and W. W. Grabowski, The multidimensional positive definite advection transport algorithm: Nonoscillatory option, *J. Comp. Phys., 86*, 355-375, 1990.

Staniforth, A., J. Côté, and J. Pudykiewicz, Comments on "Smolarkiewicz's dèformational flow", *SIAM J. Numer. Anal., 115*, 894-900, 1987.

Sweby, P. K., High resolution schemes using flux limiters for hyperbolic conservation laws, *Mon. Weather Rev., 21*, 995-1011, 1984.

Takacs, L. L., A two step scheme for the advection equation with minimal dissipation and dispersion errors, *Mon. Weather Rev., 113*, 1050-1065, 1985.

Thuburn, J., Multidimensional flux-limited advection schemes, *J. Comp. Phys., 123*, 74-83, 1996.

Thuburn, J., TVD schemes, positive schemes, and the universal limiter, *Mon. Weather Rev., 125*, 1990-1993, 1997.

Toon, O. B., R. P. Turco, D. Westphal, R. Malone and M. S. Liu, A multidimensional model for aerosols: Description of computational analogs, *J. Atmos. Sci., 45*, 2123-2132, 1988.

Tremback, C. J., J. Powell, W. R. Cotton, and R. A. Pielke, The forward-in-time upstream advection scheme: Extension to higher orders, *Mon. Weather Rev., 115*, 540-555, 1987.

van Leer, B., Toward the ultimate conservative difference scheme, IV: A new approach to numerical convection, *J. Comp. Phys., 23*, 276-299, 1977.

Walcek, C. J., and N. M. Aleksic, A simple but accurate mass conservative, peak preserving, mixing ratio bounded advection algorithm with FORTRAN code, *Atmos. Environ., 32*, 3863-3880, 1998.

Yamartino, R. J., Nonnegative conserved scalar transport using grid-cell-centered spectrally constrained Blackman cubics for applications on a variable-thickness mesh, *Mon. Weather Rev., 121*, 753-763, 1993.

C. J. Walcek, Atmospheric Sciences Research Center, State University of New York at Albany, 251 Fuller Rd., Albany, NY 12203-3649. (walcek@asrc.cestm.albany.edu)

(Received July 14, 1999; revised September 22, 1999;
accepted November 22, 1999.)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D7, PAGES 9349–9363, APRIL 16, 2000

Heat transport and weakening of atmospheric stability induced by mesoscale flows

G. A. Dalu

Cooperative Institute for Research in the Atmosphere, CIRA-CSU, Fort Collins, Colorado
Institute for Atmospheric Physics, IFA-CNR, Rome, Italy

R. A. Pielke Sr. and P. L. Vidale

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

M. Baldi

Cooperative Institute for Research in the Atmosphere, CIRA-CSU, Fort Collins, Colorado
Institute for Atmospheric Physics, IFA-CNR, Rome, Italy

Abstract. We present an analytical evaluation and interpretation on how diabatic heating of the convective boundary layer (CBL) is transported upward into the midtroposphere by mesoscale flows, and how the air mixes with the environment and therefore weakens the atmospheric static stability. The thermodynamic imprint on the free atmosphere due to the irreversible processes such as mixing, dissipation, and diffusion, associated with the mesoscale flow, is more clearly shown when the forcing is periodic in time. Convective mixing in the CBL accounts for a thermodynamic perturbation of the order of a few degrees, while mixing associated with the mesoscale activity accounts for a perturbation of the order of half a degree. To isolate this last effect, we prescribe a periodic forcing with a 1 day period, so over 24 hours, the net diabatic input averages to zero, and the contribution due to the advection cancels out. In this formulation the perturbation is solely due to irreversible processes associated with the mesoscale. These perturbations are relevant, since they are smaller, but of the same order of magnitude as perturbations associated with mesoscale advection and the CBL mixing. A more complete evaluation of the relative contribution to the atmospheric perturbations due to the mesoscale activity was completed using an initial value problem approach. In this case, there is a net transport of the diabatic heat induced by the mesoscale flow. As a consequence, when the mesoscale flow persists for several days, the static stability of the atmosphere is eroded by the combined action of the diabatic heat, CBL mixing, and transport and mixing due to the mesoscale activity. In this paper we first evaluate the contribution of the irreversible processes using a periodic in time forcing. Then we examine the atmospheric impact due to a sequence of several sea breeze days, starting from rest at time zero and letting the flow evolve as an initial value problem. Results suggest that perturbations associated with mesoscale flows generated by landscape variability are of climatological importance and need to be introduced in a parametric form in coarser large-scale models, as presently is done with turbulent subgrid CBL processes.

1. Introduction

In regions characterized by land surface heterogeneity the lower boundary generates local atmospheric flows that redistribute momentum and diabatic heat within the planetary boundary layer (PBL) [*Avissar and Schmidt*, 1998; *Shen and Leclerc*, 1995; *Dalu et al.*, 1996; *Emori*, 1998].

Integrated measurements performed during the Boreal Ecosystem-Atmosphere Study (BOREAS) experiment [*Sellers et al.*, 1995] from different platforms, such as towers, balloons, and instrumented airplanes, have shown that persistent mesoscale flows can leave an imprint in the free atmosphere which extends up to the middle of the troposphere [*Pielke and Vidale*, 1995]. One important consequence is that air perturbed by the diabatic sensible heat, instead of being confined within the PBL, is transported upward and, as it diffuses, mixes with the air in the middle of the troposphere, weakening its static stability. As a result, a few days of persistency of the mesoscale flow can add up to a signal of climatological relevance [*Pielke and Vidale*, 1995].

Mesoscale flows are driven by the horizontal gradients of sensible heat fluxes, which are usually associated with landscape variability, such as due to the presence of large bodies of water, patches of terrain covered by a different kind of vegetation, and local orographic features. When the large-scale flows are weak, the surface-forced mesoscale flows, with its distinctive diurnal pattern, can persist for several days, or even weeks. The characteristics of the air above this region are affected and modified by these mesoscale winds, till they are eventually advected away by larger-scale strong atmospheric flows. Data collected during the BOREAS experiment show sizable and persistent mesoscale atmospheric flows induced by landscape variability [*Vidale et al.*,

Copyright 2000 by the American Geophysical Union.

Paper number 1999JD901064.
0148-0227/00/1999JD901064$09.00

1997]. These flows, studied by *Vidale et al.* [1997] using a numerical model (RAMS), influence the entire lower half of the troposphere through a redistribution of heat and momentum, when the large-scale conditions are favorable.

In this paper we provide a quantitative evaluation of the intensity of the perturbations induced by mesoscale flows, and of the space and timescales involved in these processes, using linear theory. Here we analyze some of the main features induced by persistent (a few days) mesoscale flows using linear theory. While only the linear response, of course, can be evaluated using linearized equations, such models have been shown to be effective in identifying dominant physical processes associated with sea breezes [e.g., *Rotunno*, 1983; *Dalu and Pielke*, 1989, 1993; *Dalu et al.*, 1996]. This approach can also assist in the development of a parameterization of mesoscale effects generated by spatial variability of surface sensible heat and used in larger-scale models and as discussed by *Avissar and Chen* [1993], *Avissar and Schmidt* [1998], and *Wang et al.* [1998].

2. Governing Equations and Atmospheric Response

The two-dimensional and Boussinesq primitive equations in linear form are

$$\left(\frac{\partial}{\partial t}+\lambda\right)u-fv+\frac{\partial\phi}{\partial x}=K\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)u, \tag{1}$$

$$\left(\frac{\partial}{\partial t}+\lambda\right)v+fu=K\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)v, \tag{2}$$

$$\left(\frac{\partial}{\partial t}+\lambda\right)w+\frac{\partial\phi}{\partial z}-b=K\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)w, \tag{3}$$

$$\left(\frac{\partial}{\partial t}+\lambda\right)b+N_0^2w=Q+K\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)b, \tag{4}$$

$$\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}=0. \tag{5}$$

As a lower boundary condition, we assume that the vertical momentum component vanishes at the ground, $w(x, z = 0, t) = 0$. In these equations, u, v, w are the velocity components, b is the buoyancy force, Q is the diabatic buoyancy source related to the vertical divergence of the sensible heat flux in the convective boundary layer (CBL), and ϕ is the geopotential as in the work of *Rotunno* [1983]. Equation (5) is the continuity equation.

We study mesoscale flows, which are diabatically forced in the CBL, as shown, for example, by *Dalu and Pielke* [1993], and how the atmospheric stability parameters can be modified by mesoscale flows on a horizontal scale of the order of two Rossby radii. When the mesoscale flows persist one or few days in weak large-scale flow conditions, the perturbation can significantly influence the lower half of the troposphere. We use the following values for the Brunt-Väisälä frequency N_0, Coriolis parameter f, and diffusion coefficient K:

$$\Theta_z=3[\mathrm{K/km}]\quad b=g\,\frac{\theta}{\Theta},\quad N_0=\left(g\,\frac{\Theta_z}{\Theta}\right)^{1/2},$$

$$K=10\text{ to }100\ [\mathrm{m^2/s}], \tag{6}$$

$$\omega=\frac{2\pi}{\mathrm{day}}\quad f=2\omega\sin(54°)\qquad \lambda=2\text{ to }10\ [\mathrm{day^{-1}}], \tag{7}$$

where θ is the potential temperature perturbation, Θ is the environment potential temperature, and Θ_z is its vertical gradient. The parameter λ^{-1} is the lifetime of the mesoscale flow; λ is a bulk dissipation that represents frictional losses at low wavenumbers, while K is the diffusion coefficient which accounts for the losses at high wave numbers.

In the presence of periodic forcing we keep the value of the diffusion coefficient constant and equal to $K = 10$ [m²/s] and study the mesoscale response for a Rayleigh friction coefficient, $\lambda = 2$ to 10 [day^{-1}]. We conclude, from the time lag between the diabatic forcing and the mesoscale response that a realistic value of λ is closer to the higher values of the Rayleigh friction coefficient. When $\lambda = 10$ [day^{-1}], the spin-down time mesoscale flow is 2.5 hours.

The mesoscale flow is driven by the horizontal gradient of the vertical divergence of the diabatic source Q in (4):

$$Q=Q_0\,q(t)r(x,z), \tag{8}$$

where $q(t)$ is its time behavior and $r(x, z)$ its spatial distribution. The equations (1)–(5) can be reduced to an equation for the stream function ψ:

$$\left[\left(\frac{\partial}{\partial t}+\lambda-K\nabla^2\right)^2+f^2\right]\frac{\partial^2\psi}{\partial z^2}+\left[\left(\frac{\partial}{\partial t}+\lambda-K\nabla^2\right)^2+N^2\right]\frac{\partial^2\psi}{\partial x^2}=-\frac{\partial Q}{\partial x},$$

$$\nabla^2\equiv\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2},\qquad \frac{\partial\psi}{\partial z}=u,\qquad \frac{\partial\psi}{\partial x}=-w. \tag{9}$$

3. Atmospheric Response to a Periodic in Time Forcing

In this section we evaluate the thermodynamic imprint due to the irreversible processes. In order to isolate this contribution we apply a periodic forcing where the air particles describe closed orbits; as a result, the net diabatic input and the advection vanish over a cycle.

3.1. Diabatic Periodic Forcing

Here we assume that the diabatic forcing Q is periodic with a pulsation ω from which it follows that in a cycle the net diabatic input averages to zero:

$$Q=Q_0\,\mathrm{Im}\left[\frac{\exp(j\omega t)}{j}\right]\frac{1}{2}\left[1+\tanh\left(\frac{x}{L}\right)\right]\left(\frac{h-z}{h}\right)He(h-z),$$

$$Q=Q_0\sin(\omega t)\,\frac{1}{2}\left[1+\tanh\left(\frac{x}{L}\right)\right]\left(\frac{h-z}{h}\right)He(h-z),$$

$$Q_0=N_0^2h\omega,\quad h=3\ \mathrm{km},\quad L=30\text{–}100\ \mathrm{km},\quad \Theta_z=3\ \mathrm{K/km}. \tag{10}$$

$He(h - z)$ is the Heaviside step function equal to 1 when $0 < z < h$ and zero when $z > h$; h is the depth of the CBL. L, ranging from 30 to 100 km, is the horizontal extent of the influence region of the mesoscale flow, and $K = 10$ m²/s is the diffusion coefficient, and λ ranges from 2 to 10 d^{-1}.

The Fourier transform of the streamfunction equation (equation (9)), given the periodic in time forcing (equation (10)), is

$$\{[j\omega + \lambda + K(k^2 + \mu^2)]^2 + f^2\}\mu^2\tilde{\psi}$$

$$+ \{[j\omega + \lambda + K(k^2 + \mu^2)]^2 + N_0^2\}k^2\tilde{\psi} = \tilde{Q},$$

$$= Q_0 \frac{\frac{\pi}{2} kL}{\sinh\left(\frac{\pi}{2} kL\right)} \frac{2}{\mu}\left[1 - \frac{\sin(\mu h)}{\mu h}\right]. \tag{11}$$

In (11), $\tilde{\psi}$ is the Fourier transformed stream function, and k and μ are the horizontal and the vertical wave numbers, respectively. Solutions to (11), including momentum components, particle displacements, and temperature perturbations are given in Appendix A1.

3.2. Mesoscale Flow Intensity and Time Lag in Its Response to Forcing

Since the model is linear, the intensity of the flow increases monotonically as the forcing increases. The in-phase component φ_1 decreases monotonically as dissipation and diffusion increase, while the out-of-phase component φ_2 initially increases as dissipation and diffusion increase, to decrease monotonically for large values of dissipation and diffusion (equations (33) and (34) of Appendix A1). The flow components φ_1 and φ_1 are

$$\tilde{\varphi}_1 \sin(\omega t) + \tilde{\varphi}_2 \cos(\omega t)$$

$$= \frac{\tilde{Q}[(a^2\mu^2 + b^2k^2)\sin(\omega t) - c(k^2 + \mu^2)\cos(\omega t)]}{(a^2\mu^2 + b^2k^2)^2 + c^2(k^2 + \mu^2)^2}$$

$$\approx \{\tilde{Q}[(\lambda^2 + f^2 - \omega^2)\mu^2 + N_0^2k^2]\sin(\omega t)$$

$$- 2\tilde{Q}\omega\lambda(k^2 + \mu^2)\cos(\omega t)\}$$

$$\cdot\{[(\lambda^2 + f^2 - \omega^2)\mu^2 + N_0^2k^2]^2 + 4\omega^2\lambda^2(k^2 + \mu^2)^2\}^{-1} \tag{12}$$

when

$$\lambda \gg K(k^2 + \mu^2).$$

While the maximum intensity of the forcing occurs at noon (equation (10)), the maximum intensity of the mesoscale flow occurs later because of the delay induced by the dissipation. The phase lag ϑ in the response can be computed from (33) and (34) of Appendix A1. The phase lag ϑ and the time lag τ_{lag} increase as the dissipation increases, and they are small at high wave numbers and large at low wave numbers; that is, small-scale mesoscale features respond quicker to forcing:

$$\vartheta = \tan^{-1}\left[\frac{\tilde{\varphi}_1}{\tilde{\varphi}_2}\right] = \tan^{-1}\left[\frac{c(k^2 + \mu^2)}{a^2\mu^2 + b^2k^2}\right]$$

$$\approx -\tan^{-1}\left[\frac{2\omega\lambda\mu^2}{(\lambda^2 + f^2 - \omega^2)\mu^2 + N_0^2k^2}\right], \tag{13}$$

$$\tau_{\text{lag}} = \vartheta/\omega.$$

The values of the coefficients a, b, and c in (12) and (13) are given in (35) of the Appendix. The last formulas in (12) and (13) hold when the diffusion is small. We recall that a fully developed sea breeze has a horizontal scale of two Rossby radius, $k \approx 1/R_0$, and a vertical scale of double the CBL depth, $\mu \approx 1/h$, in (12) and in (13) [*Dalu and Pielke*, 1993], which gives a time lag $\tau_{\text{lag}} \approx [2\lambda/(\lambda^2 + f^2 - \omega^2) + N_0^2h^2/R_0^2]$. At high latitudes, the time lag between the forcing and the flow response is of the order of the lifetime of the flow; furthermore, the time lag and the phase shift are smaller at higher wave number, increase with dissipation, and decrease with latitude, vanishing in the absence of dissipation and diffusion.

Haurwitz [1947], analyzing several sea breeze cases, has empirically determined the Rayleigh friction, as $\lambda = 10$ [day^{-1}], as the value which induces the observed phase lag from a pure inertial oscillation in the hodograph of the horizontal wind vector. *Wang et al.* [1996] evaluate the mesoscale lifetime as λ^{-1} from 0.25 to 0.5 day, which is considerably longer than that evaluated by Haurwitz. Recently, using physical considerations, *Grant* [1997] has given as decay time $\lambda^{-1} = (u_*/h)^{-1} \approx 0.2$ day, close to Haurwitz's value.

Simulations increasing the value of the dissipation and diffusion show that the intensity of the flow decreases as the frictional force increases, while the time lag between the forcing and the mesoscale response increases for increasing values of the dissipation λ.

From (13) and from simulations the time lag τ_{lag} is from 1 to 2.5 hours for values of λ from 2 to 10 day^{-1}. Since the observed time lag between the forcing and the sea breeze response is of the order of 2–3 hours, we believe that the correct value for λ is close to that given by *Haurwitz* [1947] and by *Grant* [1997].

3.3. Thermodynamic Imprint and Air Particle Trajectories

The periodic case is not only introductory to the more complex initial value problem, but since when the forcing is periodic, the net diabatic input averages to zero in a day (equation (46)), this makes the evaluation of the net thermodynamic imprint due to irreversible processes in the free atmosphere easier to compute and clearer to understand.

In a time cycle, each air particle describes a trajectory which is a closed elongated ellipse, (Figures 1a, 2a, and 3a). The ratio between the minor and the major axis of the ellipses increases as the dissipation increases: the major axis becomes smaller as the minor axis becomes larger. Since the trajectories are closed, the net advected heat averages to zero (equation (47)). However, as the air particles describe their elliptical trajectories, they diffuse and partially mix in the environment. Each air particle interacts with the neighboring particles. The interaction distance is $r = \sqrt{K/(\lambda^2 + \omega^2)^{1/2}}$, and the timescale is $\tau = 1/\sqrt{(\lambda^2 + \omega^2)}$; see (44). The dissipative-diffusive process leaves a thermodynamic imprint with a nonzero average, (equation (48)), as shown in Figures 1b, 2b, and 3b. This imprint is of the order of tenths of degrees and may be of climatological significance, when integrated over all the coastal regions of the Earth [*Walsh*, 1974]. From (48) and (43) the temperature perturbation averaged over a cycle is

$$\overline{\delta\theta'(x, z)} = -\frac{\Theta_z}{2}\mathcal{F}^{-1}\left\{\bar{a}_1\left[(\tilde{w}_1 + \tilde{u}_1)\frac{\partial\tilde{\zeta}_1}{\partial x} + (\tilde{w}_2 + \tilde{u}_2)\frac{\partial\tilde{\zeta}_2}{\partial x}\right]\right.$$

$$\left. - \bar{a}_2\left[(\tilde{w}_2 + \tilde{u}_2)\frac{\partial\tilde{\zeta}_1}{\partial z} - (\tilde{w}_1 + \tilde{u}_1)\frac{\partial\tilde{\zeta}_2}{\partial x}\right],\right\}$$

$$\cdot\,\bar{a}_1 + j\tilde{a}_2 = \frac{[\lambda + K(k^2 + \mu^2)] - j\omega}{[\lambda + K(k^2 + \mu^2)]^2 + \omega^2}$$

$$\approx \frac{\lambda - j\omega}{\lambda^2 + \omega^2} \tag{14}$$

when

$$\lambda \gg K(k^2 + \mu^2).$$

When the dissipation and diffusion are negligible, the net thermodynamic imprint is also negligible. The imprint in-

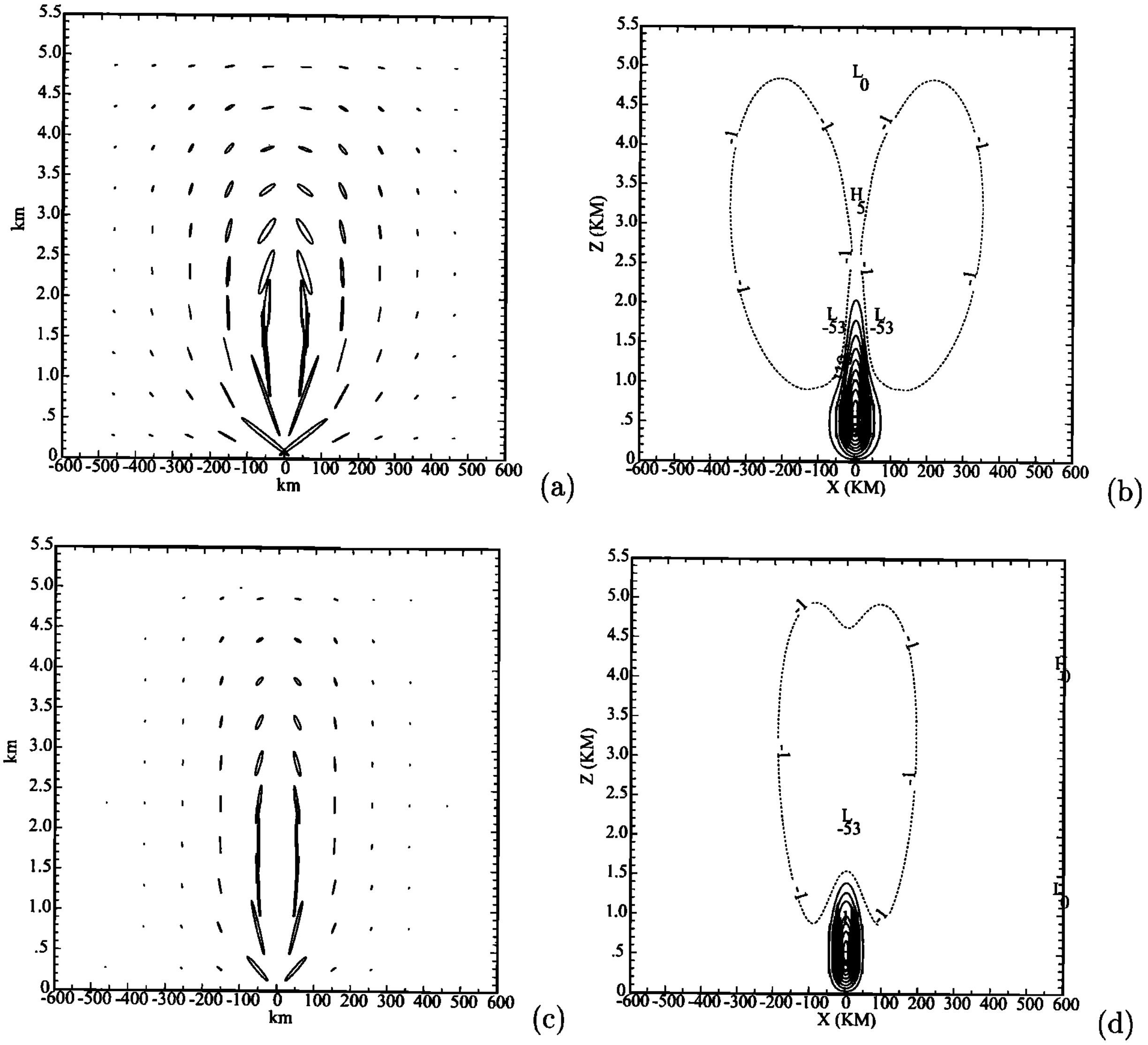


Figure 1. Air particle trajectories described for a periodic case for different values of the parameters: (a) $L_0 = 30$ km, $K = 10$ m^2/s, $\lambda = 2$ day^{-1}; and (b) averaged temperature perturbation in 1/1000 of degree. The thermodynamic imprint of the air particle trajectories is in Figure 1a. (c, d) As in Figures 1a and 1b, but $\lambda = 10$ day^{-1}.

creases with dissipation and diffusion; however, for very large values of dissipation, it decreases since the intensity of the flow decreases and the dimensions of the ellipses described by the air particles collapse. The maximum intensity of the flow and of the imprint occur when $f^2 \approx \omega^2 - [\lambda + K(k^2 + \mu^2)]^2$, which, in midlatitudes, is a value of λ close but smaller than that given by *Haurwitz* [1947] and by *Grant* [1997].

4. Diabatic Forcing in the Initial Value Problem

In this section we evaluate the thermodynamic tropospheric imprint of the mesoscale flow. We use an initial value problem approach. With a start from a rest state, we apply the diabatic forcing for four consecutive days. We evaluate the perturbations induced by the diabatic input, the advection, and the irreversible processes combined. In this section the Rayleigh friction coefficient is set to $\lambda = 5$ [day^{-1}], which corresponds to a spin-down time of the order of 5 hours in the absence of forcing. The diffusion coefficient ranges from $K = 10$ [m^2/s] to $K = 100$ [m^2/s]; larger values refer to the CBL region.

4.1. Diabatic Forcing

In the initial value problem, the atmosphere is initially unperturbed and uniformly stratified. The diabatic forcing is

$$Q(x, z, t) = Q_0 He(h - z)\,\frac{1}{2}\left[1 + \tanh\left(\frac{x}{L_0}\right)\right] \cdot [A + \sin(\omega t)], \qquad (15)$$

$$Q_0 = N_0^2 \omega h_0, \qquad L_0 = 30 \text{ km}.$$

N_0 is the Brunt-Väisälä frequency and $\omega h_0 = 2$ km/d is the growth rate of the CBL. $Q > 0$ during the daytime ($t_{\text{sunrise}} < t < t_{\text{sunset}}$), and $Q < 0$ during the night ($t_{\text{sunset}} < t < t_{\text{sunrise}}$). When $A = 0$, the duration of the day is equal to the length of the night. We chose A such as to have a daytime of 14 hours

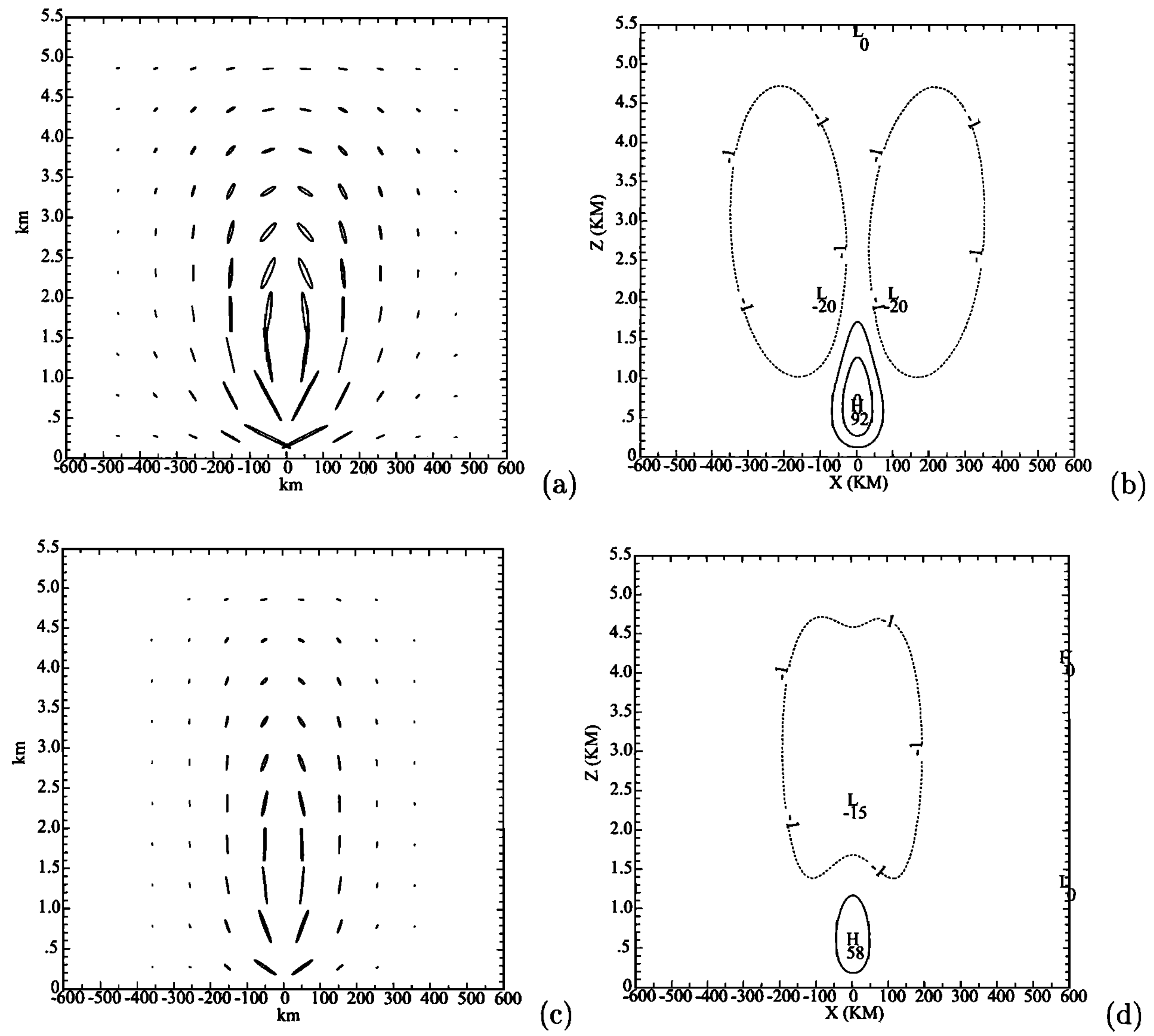


Figure 2. (a) Air particle trajectories with L_0 = 60 km, K = 10 m^2/s, λ = 2 day^{-1}; and (b) averaged temperature perturbation. The imprint of the air particle trajectories is in Figure 2a. (c, d) As in Figures 2a and 2b, but λ = 10 day^{-1}.

and a nighttime of 10 hours, as during a summer day at 54°N latitude.

During the first day ($t_{\text{sunrise}} < t < t_{\text{sunset}}$) the CBL, $h_D(x, t)$, grows monotonically from sunrise to sunset [*Green and Dalu*, 1980]:

$$h_D(x, t) = \frac{2}{N_0^2}\int_{t_{\text{sunrise}}}^{t} dt'\, Q = \frac{1}{2}h_{0_D}(t)\left[1 + \tanh\left(\frac{x}{L_0}\right)\right],$$

$$h_{0_D}(t) = \frac{2Q_0}{N_0^2}\left[At + \frac{1 - \cos(\omega t)}{\omega}\right]. \tag{16}$$

The diabatic heat is distributed through the CBL as

$$Q_D(x, z, t) = Q_0 He(h_{0_D}(t) - z)\,\frac{1}{2}\left[1 + \tanh\left(\frac{x}{L_0}\right)\right] \cdot [A + \sin(\omega t)]. \tag{17}$$

During the following night ($t_{\text{sunset}} < t < t_{\text{sunrise}}$) the air is cooled through a shallower layer, $h_N(x, t)$:

$$h_N(x, t) = h_{0_N}(t)\,\frac{1}{2}\left[1 + \tanh\left(\frac{x}{L_0}\right)\right], \qquad h_{0_N}(t) = 300 \text{ m}. \tag{18}$$

The diabatic cooling is distributed through the nocturnal boundary layer (NBL) as

$$Q_N(x, z, t) = Q_N He(h_{0_N}(t) - z)\,\frac{1}{2}\left[1 + \tanh\left(\frac{x}{L_0}\right)\right] \cdot [A + \sin(\omega t)]. \tag{19}$$

In the following days after the first sea breeze day, in the early hours of the morning, the convection has to work its way through the cold air of the nocturnal NBL and through the stable marine air advected inland during the night, before it can grow deeper than the CBL of the previous day. Since the depth of the density current associated with the sea breeze front is about half of the depth of the CBL of the previous day or less, this delay is $\Delta t_1 \approx$ 5–6 hours, i.e., about half of the growth time of the CBL in the first day.

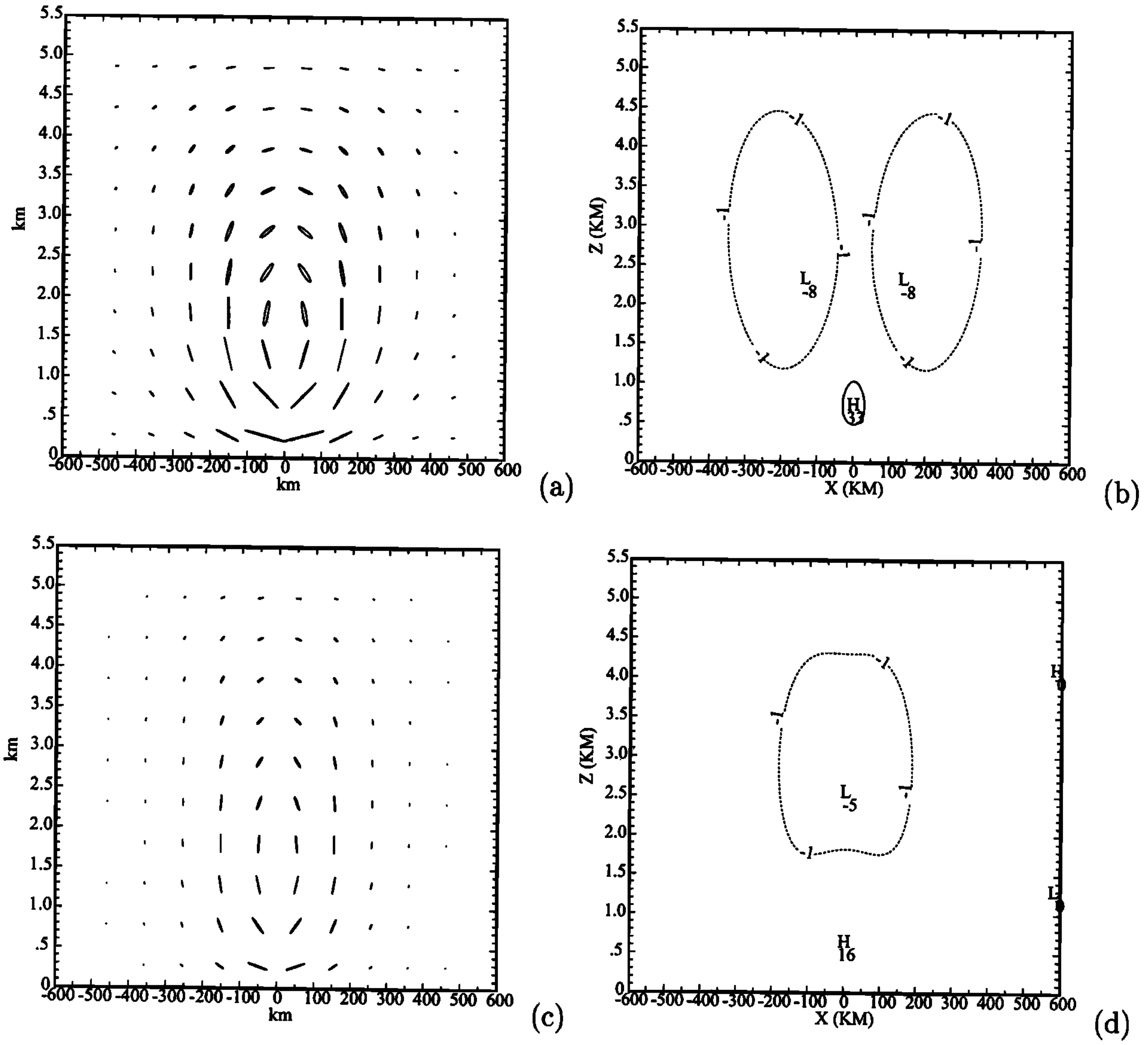


Figure 3. (a) Air particle trajectories with $L_0 = 100$ km, $K = 10$ m²/s, $\lambda = 2$ day^{-1}, and (b) averaged temperature perturbation. The imprint of the air particle trajectories is in Figure 3a. (c, d) As in Figures 3a and 3b, but $\lambda = 10$ day^{-1}.

The behavior of the diurnal CBL, $h_{0_D}(t)$, and of the nocturnal NBL, $h_{0_N}(t)$, is shown in Figure 4, for a 4 day integration.

4.2. Stream Function and Dynamical Fields

The Fourier transform of (9), Laplace transformed in time and as forced by (15), is

$$\{[s + \lambda + K(k^2 + \mu^2)]^2 + f^2\}\mu^2\hat{\psi} + \{[s + \lambda + K(k^2 + \mu^2)]^2$$

$$+ N_0^2\}k^2\hat{\psi} = Q_0 q(s) \frac{\frac{\pi}{2} kL_0}{\sinh\left(\frac{\pi}{2} kL_0\right)} \frac{2}{\mu} [1 - \cos(\mu h)], \qquad (20)$$

where $\hat{\psi}$ is the stream function, and s is the Laplace transformed time t. In (20), $h = h_{0_D}(t)$ during the day, and $h = h_{0_N}(t)$ during the night. The flow driven by the diabatic input is governed by (20). The solution to the stream function equation, the momentum components, and particle displacements are given in Appendix A2.

5. Analysis of Processes in the Initial Value Problem

Taking advantage of the additivity of the solutions, because of our linear approach to the problem, here we examine the contribution of each of the physical processes in shaping the mesoscale flow.

5.1. Daytime Diabatically Driven Flow, Nighttime Diabatically Driven Flow, and Sea Breeze Fossil Flow

In daytime the flow is driven by the diabatic warming (equation (17)). The stream function and the temperature field are shown in Figure 5, at sunset of the first day (1900 LTC), which is at the end of the diurnal warming period. At this time, the sea breeze flow shows downwelling over the sea with adiabatic warming and upwelling over the land side with a lifting of the